DOCUMENT RESUME

ED 436 087 FL 026 058

AUTHOR	Thomas, Wayne P.; Collier, Virginia
TITLE	School Effectiveness for Language Minority Students. NCBE Resource Collection Series, No. 9.
INSTITUTION	National Clearinghouse for Bilingual Education, Washington, DC.
SPONS AGENCY	Office of Bilingual Education and Minority Languages Affairs (ED), Washington, DC.
PUB DATE	1997-12-00
NOTE	97p.
CONTRACT	T295005001
AVAILABLE FROM	National Clearinghouse for Bilingual Education, George Washington University, Center for the Study of Language and Education, 2011 Eye Street, N.W., Suite 200, Washington, DC 20006. Tel: 202-467-0867.
PUB TYPE	ERIC Publications (071) -- Reports - Research (143)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Academic Achievement; Bilingual Students; Elementary Secondary Education; English (Second Language); *Language Minorities; *Limited English Speaking; Public Schools; School Effectiveness; *Second Language Learning; Urban Schools

ABSTRACT

This publication presents a summary of an ongoing collaborative research study that is both national in scope and practical for immediate local decision making in schools. It is written for bilingual and English as a Second Language program coordinators and local school policy makers. The research includes findings from five large urban and suburban school districts in various regions of the United States where large numbers of language minority students attend public schools, with over 700,000 language minority student records collected from 1982-1996. A developmental model of language acquisition for school is explained and validated by the data analyses. The model and findings from this study make predictions about long-term student achievement as a result of a variety of instructional practices. Instructions are provided for replicating this study and validating the findings in local school systems. General policy recommendations and specific action recommendations are provided for decision makers in schools. (Contains 71 references.) (Author/SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 436 087

NCBE

RESOURCE
COLLECTION
SERIES

9

December
1997

School Effectiveness for Language Minority Students

Wayne P. Thomas and Virginia Collier
George Mason University

U.S. DEPARTMENT OF EDUCATION
[illegible]
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
[illegible]

Disseminated by
National Clearinghouse for Bilingual Education
The George Washington University
Center for the Study of Language and Education
1118 22nd Street, NW
Washington, DC 20037

2

BEST COPY AVAILABLE

The National Clearinghouse for Bilingual Education (NCBE) is funded by the U.S. Department of Education's Office of Bilingual Education and Minority Languages Affairs (OBEMLA) and is operated under Contract No. T295005001 by The George Washington University, Center for Education Policy Studies/Institute for the Study of Language and Education. The contents of this publication are reprinted from the NCBE Resource Collection. Materials from the Resource Collection are reprinted "as is." NCBE assumes no editorial or stylistic responsibility for these documents. The view expressed do not necessarily reflect the views or policies of The George Washington University or the U.S. Department of Education. The mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. government. Readers are free to duplicate and use these materials in keeping with accepted publication standards. NCBE requests that proper credit be given in the event of reproduction. (v. 1.2)

BEST COPY AVAILABLE

CONTENTS

Executive Summary 6
Abstract 11

I. Urgent Needs 12

II. Overview of This Study for Decision-Makers 14
- A. The long-term picture 14
- B. Key findings of this study 15
- C. Study designed to answer urgent school policy questions 16

III. Development of This Study 18
- A. Limitations of typical short-term program evaluations 18
- B. Common misconceptions of "scientific"research in education 19
  - 1. Research questions on effectiveness 19
  - 2. Research methodology in effectiveness studies 20
    - a. Inappropriate use of random assignment 20
    - b. Statistical conclusion validity 21
    - c. External validity 23
    - d. Other internal validity concerns 24
  - 3. Research reviews on program effectiveness in LM education 24
- C. Analyzing program effectiveness in our study 26
- D. Magnitude of our study 30

IV. Our Findings: The "How Long" Research 32
- A. How long: Schooling only in L2 32
- B. How long: Schooling in both L1 and L2 35
- C. Summary of "How long" findings for English Language Learners 36
- D. How long: Bilingual schooling for native-English speakers 36
- E. How long: Influence of student background variables 37
  - 1. Proficiency in L1 and L2 37
  - 2. Age 37
  - 3. Student's first language (L1) 38
  - 4. Socioeconomic status 38
  - 5. Formal schooling in L1 39

V. **Understanding Our "How Long" Findings: The Prism Model** 40
- A. The instructional situation for the native-English speaker 40
- B. The Prism Model 42
  - 1. Sociocultural processes 42
  - 2. Language development 43
  - 3. Academic development 43
  - 4. Cognitive development 43
  - 5. Interdependence of the four components 44
- C. The instructional situation for the English Language Learner in an English-only program 44

VI. **Our Findings: School Effectiveness** 48

- A. Characteristics of effective programs 48
  - 1. L1 instruction 48
  - 2. L2 instruction 49
  - 3. Interactive, discovery learning & other current approaches to teaching 50
  - 4. Sociocultural support 51
  - 5. Integration with the mainstream 51
- B. Language minority students' academic achievement patterns 52
  - 1. The influence of elementary school bilingual/ESL programs on ELLs'achievement 53
    - a. Amount of L1 support 56
    - b. Type of L2 support 59
    - c. Type of teaching style 60
    - d. Sociocultural support 61
    - e. Integration with the curricular mainstream 62
    - f. Interaction of the five program variables 64
  - 2. The influence of secondary school ESL programs on ELLs' achievement 65
  - 3. School leavers 68

VII. **Phase II of This Study** 69

VIII. **Recommendations** 69
- A. Policy recommendations 71
- B. How is your school system doing? -- The Thomas-Collier Test 74
- C. If you failed the Thomas-Collier Test 75
- D. Action recommendations 77
- E. A Call to action 79

IX. **Endnotes** 81

X. **Appendix A -- Percentiles and Normal Curve Equivalents (NCEs)** 83

XI. **Appendix B -- Phase II of Thomas and Collier Research, 1996-2001** 88

XII. **References** 90

XIII. **About the Authors** 96

Executive Summary

This report is a summary of a series of investigations of the fate of language minority students in five large school systems during the years 1982-1996. It is different from typical existing research studies in a number of important ways. Specifically, our work:

- **is macroscopic rather than microscopic in purview.** Our research investigates the "big picture" surrounding the effects of school district instructional strategies on the long-term achievement of language-minority students in five large school districts in geographically dispersed areas of the U.S.

- **is non-interventionist rather than interventionist in philosophy.** This research avoids laboratory-style research methods (e.g. random assignment) that are inappropriate or impossible to use in typical school settings. Instead, it uses alternative and more appropriate methods of achieving acceptable internal validity (e.g., sample restriction, blocking, time-series analyses, and analysis of covariance, where appropriate). In particular, only instructional programs that are well-implemented are examined for their long-term success, in order to reduce the confounding effects of implementation differences on instructional effectiveness.

- **collects and analyzes individual student-level data (rather than summarizing existing analyses or school-and-district-wide reports)** on student characteristics, the instructional interventions they received, and the test results that they achieved years after participating in programs for language-minority students.

- **is a summary of findings from a series of quantitative case studies in each participating school district.** In each school district, researchers and school staff collaboratively analyzed a large series of "data views" that focused on questions of concern to the local school district and to the researchers. This report provides conclusions and interpretations that are robustly supported in case studies from all five school districts rather than results that are unique to one district, one set of conditions, or small, isolated groups of students.

- **emphasizes a wide range of statistical conclusion validity, external validity, and internal validity issues**, not just a few selected aspects of internal validity as in the case of many so-called "scientific" studies in this field.

- **investigates very large samples of students** (a total of more than 700,000 student records) rather than classroom-sized samples. We have collected and analyzed large sets of individual student records from a variety of offices and sources within each school district and have linked these records together at specified points in time (cross-sectional studies) and have followed large groups of students across time (longitudinal studies).

- **is built on an emergent model of language acquisition for school (Collier's Prism Model)** and further develops the interpretation of this model. In addition, the data analyses test the predictive success of this model and provide information on which variables are most important and most powerful in influencing the long-term achievement of English learners (also referred to as LEP students or ESL students).

- **provides a long-term outlook** (rather than a short-term view) for the required long-term processes necessary for English learners to reach full parity with native-English speakers. Our research emphasizes longitudinal data analyses rather than only short-term, cross-sectional, 1-2 year program evaluations as in most other research in this field.

- **emphasizes student achievement across the curriculum, not just English proficiency.** Previous research has largely ignored the fact that English learners quickly fall behind the constantly advancing native-English speakers in other school subjects (e.g., social studies, science, mathematics) during each year that the instructional program for English learners focuses mostly or exclusively on English proficiency, or offers "watered-down" instruction in other school subjects, or offers English-only instruction that is poorly comprehended by the English learners.

- **adopts the educational standards and goals for language minority students from *Castañeda v. Pickard (1981).*** This federal court case provided guidelines that school districts should select educational programs of theoretical value for English learners, implement them well, and then follow the long-term school progress of these students to assure equal educational opportunity. The researchers propose the Thomas-Collier Test as a means for school districts to self-assess their success in providing long-term equality of educational opportunity for English learners.

- **defines "success" as "English learners reaching eventual full educational parity with native-English speakers in all school content subjects (not just in English proficiency) after a period of at least 5-6 years."** A "successful educational program" is a program whose typical students reach long-term parity with national native-English speakers (50th percentile or 50th NCE on nationally standardized tests) or whose local English learners reach the average achievement level of native-English speaking students in the local school system. A "good program" is one whose typical English learners close the on-grade-level achievement gap with native-English-speaking students at the rate of 5 NCEs (equivalent to about one-fourth of a national standard deviation) per year for 5-6 consecutive years and thereafter gain in all school subjects at the same levels as native-English speaking students.

- **utilizes data mining techniques as well as quasi-experimental research techniques.** The study incorporates available student-level information in each school district with information collected by school district staff specifically for these studies.

- **consists of collaborative, participatory, and interactive investigations conducted jointly with the staff of participating school systems** who acted as joint researchers in granting access to their existing data, collecting additional data to support extended research inquiry, providing contextual understanding of preliminary findings, and providing priorities and structure for sustained investigations.

- **emphasizes action-oriented and decision-oriented research rather than conclusion-oriented research.** Our investigations are designed to diagnose the past and present situations for language minority students in participating school districts and to make formative recommendations for each school system's activities in planned reform and improvement of their programs and instruction. For maximum understanding and decision-making utility for school personnel, our quantitative findings, including measures of central tendency and variability, are presented in text, charts, and graphics rather than in extensive tables of statistics. Our discussions of instructional effect size are conservatively stated in terms of national standard deviations rather than the typically smaller local standard deviations that would lead to spuriously large effect size estimates. In addition, our recommendations are based on robust findings sustained across all of our participating school systems, increasing their generalizability and worth for local decision-making.

- **provides school personnel with data on the long-term effects of their past and present programmatic decisions** on the achievement and school success of language minority students. In addition, our work engages the participating school systems in a process of ongoing reform over the next 5-10 years.

- **strongly emphasizes the need for wide replication of our findings.** Although our findings are conclusive for our participating school districts, we strongly recommend that our research should be repeated in many more school districts and in a broader set of instructional contexts to achieve even wider generalizability. We encourage school districts to replicate our research by examining your own local long-term data. If it is not feasible to replicate our research in full, we strongly recommend that every school system conduct the abbreviated analysis described herein (the Thomas-Collier Test) in order to perform a needs assessment of your own programs for language minority students.

- **contains both educational and research re-definition components.** We describe the great limitations of past research in this field, especially that based on short-term studies with small samples or on research summaries that are based on the "vote-counting" method and not based on cumulative statistical significance or on effect size. We describe why more than 25 years of past research has not yielded useful decision-making information for use by school personnel and make suggestions for researchers who wish to produce research that is more useful to school staff. Also, we provide explanations for aspects of our methodology (e.g., the use of normal curve equivalents [NCEs] rather than percentiles or grade-equivalent scores) that we hope will be adopted by schools and researchers alike.

- **provides a theoretical foundation and a basis for continued development for our nationwide research during the next 5-10 years** that we hope will be emulated and replicated by many school districts and researchers nationwide.

In summary, we intend our research to redefine and reform the nature of research conducted for the benefit of language minority students. We propose that all future research on instructional effectiveness in this field emphasize long-term, longitudinal analyses with associated measures of effect size as well as shorter-term, cross-sectional analyses; we propose that the definition of school success for language minority students be changed to fit the "long-term parity" criteria implicit in *Castañeda v. Pickard*; and we propose that student achievement in all areas of the school curriculum be substituted for English proficiency as the primary educational outcome of programs for language minority students.

Finally, we propose the Prism model as a means of understanding how the vast majority of English learners fail in the long term to close the initial achievement gap in all school subjects with age-comparable native-English speakers. Our findings indicate that those English learners who experience well-implemented versions of the most common education programs for English learners in their elementary years, including those who spend five years or more in U.S. schools, finish their school years at average achievement levels between the 10th and 30th national percentiles (depending on the type of instruction received) when compared to native-English-speaking students who typically finish school at the 50th percentile nationwide. In particular, our findings indicate that students who receive well-implemented ESL-pullout instruction, a very common program nationwide, and then receive years of instruction in the English mainstream, typically finish school with average scores between the 10th-18th national percentiles, or do not even complete high school. In contrast, English learners who receive one of several forms of enrichment bilingual education finish their schooling with average scores that reach or exceed the 50th national percentile.

We point out that these findings constitute a wake-up call to U.S. school systems and should underscore the importance of the need for every school district to conduct its own investigation to examine the long-term effects of its existing programs for English learners. If our national findings are confirmed in a school district as a result of the local investigation, and we believe that they will be, then wholesale review and reform of local instructional strategies for English learners as well as all language minority students are in order. We propose the Prism Model, as further developed and tested by these data analyses, as a theoretical basis for improving existing instructional strategies, and for developing new ones to meet the assessed long-term needs of English learners. These instructional strategies are the key to demonstrably helping our substantially increasing numbers of language minority students to reach adulthood as fully functional and productive U.S. citizens who will be able to sustain our current favorable economic climate well into the 21st century. We solicit the participation and assistance of researchers and school districts nationwide to address these most urgent educational issues.

ABSTRACT

This publication presents a summary of an ongoing collaborative research study that is both national in scope and practical for immediate local decision-making in schools. This summary is written for bilingual and ESL program coordinators, as well as for local school policy makers. The research includes findings from five large urban and suburban school districts in various regions of the United States where large numbers of language minority students attend public schools, with over 700,000 language minority student records collected from 1982-1996. A developmental model of language acquisition for school is explained and validated by the data analyses. The model and the findings from this study make predictions about long-term student achievement as a result of a variety of instructional practices. Instructions are provided for replicating this study and validating these findings in local school systems. General policy recommendations and specific action recommendations are provided for decision makers in schools.

URGENT NEEDS

During the past 34 years in the United States, the growing and maturing field of bilingual/ESL education experienced extensive political support in its early years, followed by periodic acerbic policy battles at federal, state, and local levels in more recent years. Too often the field has remained marginalized in the eyes of the education mainstream. Yet over these same three decades, a body of research theory and knowledge on schooling in bilingual contexts has gradually expanded the field's conception of effective schooling for culturally and linguistically diverse school populations. Unfortunately, this emerging understanding has been clouded by those who have insisted on short-term investigations of complex, long-term phenomena, and by those who have mixed studies of stable, well-implemented instructional programs with evaluations of unstable, newly-created programs. The available knowledge from three decades of research has also been obscured by those who insist on describing programs as either "bilingual" or "English-only," completely ignoring the fact that some forms of bilingual education are much more efficacious than others, and that the same is true for English-only programs. What we've learned from research has not been put into practice by those decision-makers at the federal, state, and local levels who determine the nature of educational experiences that language minority students receive. These students, both those proficient in English and those just beginning to acquire English, have traditionally been under-served by U.S. schools.

As federal funding for education varies from year to year, state and local governments remain heavily responsible for meeting student needs, both for language minority students, and for those who are part of the English-speaking majority. But local and state decision-makers have had little or no guidance and have, by necessity, made instructional program decisions based on their professional intuition and their personal experience, frequently in response to highly politicized input from special interest groups of all sorts of persuasions. **What has been needed, and what this research provides, is a data-based (rather than opinion-based) set of instructional recommendations that tell state and local education decision-makers what will happen in the long-term to language minority students as a result of their programmatic decisions made now.**

Why is this such an urgent issue? U.S. demographic changes demand this reexamination of what we are doing in schools. In 1988, 70 percent of U.S. school-age children were of Euro-American, non-Hispanic background. But by the year 2020, U.S. demographic projections predict that at least 50 percent of school-age children will be of non-Euro-American background (Berliner & Biddle, 1995). By the year 2030, language minority students (approximately 40 percent), along with African-American students (approximately 12-15 percent), will be the majority in U.S. schools. By the year 2050, the total U.S. population will have doubled from its present levels, with approximately one-third of the increase attributed to immigration (Branigin, 1996). Since non-Euro-American-background students have generally not been well served by our traditional forms of education during most of the 20th century, and since the percentage of school-age children in this under served category will increase dramatically in the next quarter-century, many schools are now beginning to reexamine their instructional and administrative practices, to find better ways to serve all students.

Also, the urgency for changes in schooling practices is driven by current U.S. patterns of high school completion. In school policy debates regarding provision of special services for new arrivals from other countries, someone often mentions a family member who emigrated to the U.S. in the first half of the 20th century, received no special services, and "did just fine." But half a century ago, a high school diploma was not needed to succeed in the work world, with only 20 percent of the U.S. adult population having completed high school as of 1940. Half a century later in 1993, 87 percent of all adults in the U.S. have completed at least a high school education, and 20 percent of the total have also completed a four-year college degree or more (National Education Goals Panel, 1994). The modern world is much more educationally competitive than the world of 50 years ago. Those who were able to "do just fine" with less-than-high-school education 50 years ago would face much more formidable challenges now, as the minimum- necessary education for good jobs and for productive lives has greatly increased. This trend will only accelerate in the next 25 years.

Thus as we face the 21st century, effective formal schooling has become an essential credential for **all** adults to compete in the marketplace, for low-income as well as middle-income jobs. Just to put food on the table for one's family, formal schooling is crucial, and successful high school completion is the minimum necessary for a good job and a rewarding career. **Schooling must thus be made accessible, meaningful, and effective for all students, lest we create an under-educated, under-employed generation of young adults in the early 21st century.** The research findings of the studies presented in this publication demonstrate that we can improve the long-term academic achievement of language minority students, our schools' fastest growing group. By reforming current school practices, all students will enjoy a better educated, more productive future, for the benefit of all American citizens who will live in the world of the next 15-25 years. It is in the self-interest of all citizens that the next adult generations be educated to meet the enormously increased educational demands of the fast-emerging society of the near future.

OVERVIEW OF THIS STUDY FOR DECISION-MAKERS

We designed this study to address educators' immediate needs in decision-making. We wanted to provide a national view of language minority students across the U.S. by examining who they are and what types of school services are provided for them. We then linked student achievement outcomes to the student and instructional data, to examine what factors most strongly influence these students' academic success over time. When examining the data and collaboratively interpreting the results with school staff in each of our five school district sites, **we have discovered consistent patterns across school districts that are very generalizable beyond the individual school contexts in which each study has been conducted. In this publication, we are reporting these generalizable patterns.**

The Long-Term Picture

One very clear conclusion that has emerged from the data analyses in our study is the importance of gathering data over a long period of time. We have found that examination of language minority students' achievement over a 1-4 year period is too short-term and leads to an inaccurate perception of students' actual long-term performance, especially when these short-term studies are conducted in the early years of school. Thus, we have focused on gathering data across all the grades K-12, with academic achievement data in the last years of high school serving as the most important measures of academic success in our study. Many studies of school effectiveness as well as program evaluations in bilingual/ESL education have focused on the short-term picture for funding and policy purposes, examining differences between programs in the early grades, K-3. In our current research, we have found data patterns similar to those often reported in other short-term studies focused on Grades K-3--little difference between programs. Thus, those who say that there is little or no difference in student achievement across programs (e.g., ESL pullout vs. transitional bilingual education, for example) are quite correct if one only examines short-term student data from the early grades. However, **significant differences in program effects become cumulatively larger, and thus more apparent, as students continue their schooling in the English-speaking mainstream (grade-level classes).[1] Only those groups of language minority students who have received strong cognitive and academic development through their first language for many years (at least through Grade 5 or 6), as well as through the second language (English), are doing well in school as they reach the last of the high school years.**

Thus, the short-term research does not tell school policy makers what they need to know. They need to know what instructional approaches help language minority students make the gains they need to make AND CONTINUE TO SUSTAIN THE GAINS throughout their schooling, especially in the secondary years as instruction becomes cognitively more difficult and as the content of instruction becomes more academic and abstract. We have found that only quality, long-term, enrichment bilingual programs using current approaches to teaching, such as one-way and two-way developmental bilingual education,[2] when implemented to their full potential, will give language minority students the grade-level cognitive and academic development

BEST COPY AVAILABLE

needed to be academically successful in English, and to sustain their success as they reach their high school years. We note that many bilingual programs and many English-only programs fail to meet these standards. In addition, we have found that some types of bilingual programs are no more successful than the best English-only programs in the long term.

Many English learners receive instructional programs that are too short-term in focus, or fail to provide consistent cognitive development in students' first language, or allow students to fall behind their English-speaking peers in other school subjects while they are learning English, or are not cognitively and academically challenging, or are poorly implemented. These programs typically fail to help students sustain their early achievement gains throughout their schooling, especially during the cognitively difficult and academically demanding years after elementary school. And **the key to high school completion is students' consistent gains in all subject areas (not just in English) with each year of school, sustained over the long term.**

Key Findings of This Study

We have found that three key predictors of academic success appear to be more important than any other sets of variables. These school-influenced factors can be more powerful than student background variables or the regional or community context. For example, these school predictors have the power to overcome factors such as poverty at home, or a school's location in an economically depressed region or neighborhood, or a regional context where an ethnolinguistic group has traditionally been underserved by U.S. schools. Schools that incorporate all three of the predictors discussed below are likely to graduate language minority students who are very successful academically in high school and higher education.

The first predictor of long-term school success is cognitively complex on-grade-level academic instruction through students' first language for as long as possible (at least through Grade 5 or 6) **and cognitively complex on-grade-level academic instruction through the second language (English) for part of the school day,** in each succeeding grade throughout students' schooling. Here, we define students' first language as the language in which the child was nursed as an infant. Children raised bilingually from birth benefit strongly from on-grade-level academic work through their two languages, as do children dominant in English who are losing their heritage language. Children who are proficient in a language other than English and are just beginning development of the English language when they enroll in a U.S. school benefit from on-grade-level work in two languages as well. In addition, English-speaking parents who choose to enroll their children in two-way bilingual classes have discovered that their children also benefit strongly from academic work through two languages. In our research, we have found that children in well-implemented one-way and two-way bilingual classes outperform their counterparts being schooled in well-implemented monolingual classes, as they reach the upper grades of elementary school. Even more importantly, they sustain the gains they have made throughout the remainder of their schooling in middle and high school, even when the program does not continue beyond the elementary school years.

The second predictor of long-term school success is the use of current approaches to teaching the academic curriculum through two languages. Teachers and students are partners in discovery learning in these very interactive classes that often use cooperative learning strategies for group work. Thematic units help students explore the interdisciplinary nature of problem-solving through cognitively complex, on-grade-level tasks, incorporating technology, fine arts, and other stimuli for tapping what Gardner (1993) calls the "multiple intelligences." The curriculum reflects the diversity of students' life experiences across sociocultural contexts both in and outside the U.S., examining human problem-solving from a global perspective. Language and academic content are acquired simultaneously, with oral and written language viewed as an ongoing developmental process. Academic tasks directly relate to students' personal experiences and to the world outside the school.

The third predictor is a transformed sociocultural context for language minority students' schooling. Here, the instructional goal is to create for the English learner the same type of supportive sociocultural context for learning in two languages that the monolingual native-English-speaker enjoys for learning in English. When school systems succeed at this, they create an additive bilingual context,[3] and additive bilingual contexts are associated with superior school achievement around the world. For example, an additive bilingual context can be created within a school with supportive bilingual staff, even in a region of the U.S. where subtractive bilingualism is prevalent. One way that some schools have transformed the sociocultural context for language minority students is to develop two-way bilingual classes. When native-English-speaking children participate in the bilingual classes, language minority students are no longer segregated for any portion of the school day. With time, these classes come to be perceived by the school community as what they really are--**enrichment**--rather than remedial classes. In some two-way bilingual schools with prior reputations as violent inner city schools, the community now perceives the bilingual school as the "gifted and talented" school. Changes in the sociocultural context of schooling cannot happen easily and quickly, but with thoughtful, steady changes being nurtured by school staff and students, the school climate can be transformed into a warm, safe, supportive learning environment that can foster improved achievement for all students in the long term.

Study Designed to Answer Urgent School Policy Questions

Our research has followed language minority students across time by examining a wide variety of experienced, well-managed, and well-implemented school programs that utilize different degrees of validated instructional and administrative approaches for language minority students. **At the end of the students' schooling, this research seeks to answer these questions:**

- **How much time is needed for language minority students who are English language learners to reach and sustain on-grade-level achievement in their second language?**
- **Which student, program, and instructional variables strongly affect the long-term academic achievement of language minority students?**

To address these questions, we have focused our attention on the local education level, where the educational "action" is. We have examined what exists in local school systems around the country without making any changes in the school services provided for language minority students. We have worked collaboratively with local school staff in each school district to collect long-term language minority achievement and program participation data for all Grades K-12. We have analyzed this data, have collaboratively discussed and interpreted the findings with the decision-makers in the participating school systems, and have jointly arrived at recommendations that proceed from our findings. The recommendations have led to administrative and instructional action in each school system. **In replicating this research in school systems around the country, we have achieved a body of consistent findings that we believe deserves the critical attention of school decision-makers in all states.** This report presents these findings to education policy makers, with recommendations for instructional decisions for language minority students in all U.S. school contexts.

DEVELOPMENT OF THIS STUDY

When we first conceptualized this study, the research design grew out of the research knowledge base that has developed over the past three decades in education, linguistics, and the social sciences. As we watched the field of language minority education expand its range of services to assist linguistically and culturally diverse students, we were acutely aware that little progress was being made in studies of program effectiveness for these students. Since measuring program effectiveness is an area of great concern to school administrators and policy makers, it seemed increasingly important that we address some of the flaws inherent in reliance on program evaluation data as the main measures of program effectiveness.

Limitations of Typical Program Evaluations

One of the limitations of typical program evaluations is the focus on a short-term horizon. Since once-a-year reports are often required by funding sources at state and federal levels, evaluation reports typically examine the students who happen to attend a school in a given year and are assigned to special instructional services, by comparing each student's performance on academic measures in September to that same student's performance in April or May. This is important information for teachers, who expect each student to demonstrate cognitive and academic growth with each school year. But this is not sufficient decision-making information for the administrator, who is concerned about the larger picture. The larger picture includes the diagnostic information regarding the growth each student has made in one school year, but administrators also need to know how similar students (groups of students with the same general background characteristics) are doing in each of the different services being provided, to compare different instructional approaches and administrative structures. Also administrators need to know how all groups of students do in the long term, as they move on through the program being evaluated, and continue their years in school. Program evaluators are rarely able to provide this long-term picture.

A second limitation is that students come and go, sometimes at surprising rates of mobility, making it difficult to follow the same students across a long period of time, for a longitudinal view of the program's apparent effects on students. In many school systems, those students who stay in the same school for a period of 4-6 years represent a small percentage of the total students served by the program during those years. Third, programs vary greatly in how they are implemented from classroom to classroom and from school to school, making it difficult to compare one program to another. Fourth, pretest scores in short-term evaluations typically underestimate English learners' true scores until students learn enough English to demonstrate what they really know. As a result of these limitations, administrators tend to make decisions based on the short-term picture from the data in their 1-3 years of annual program evaluation reports which normally don't provide longitudinal data. Administrators rely on the teachers' assurance that students are making the best progress that they can and take the politically expedient route with school board members and central office administrators. Given the many limitations of short-term evaluations, we have approached this study from a different perspective, to overcome some of the inherent problems in program evaluations. But before we present the research design of this study, it is also important to examine common misconceptions of research methodology in education that can lead to inaccurate reporting of research findings on program effectiveness.

Common Misconceptions of "Scientific" Research in Education

In addition to the inherent limitations for decision-making of short-term program evaluations performed on small groups of students, there are the enormous limitations of education research that is labeled "scientific" by some of its proponents (e.g. Rossell & Baker, 1996). We write this section to dispel the myths that abound in the politically-driven publications on language minority education regarding what constitutes sound research methodology for decision-making purposes. We ask that educators in this field become more knowledgeable on research methodology issues, so that language minority students do not suffer because of the misconceptions that shifting political winds stir up from moment to moment. The misinformation that is disseminated through use of the term "scientific" must be dispelled. In this section, we examine two major types of misconceptions--asking the wrong research questions, and using or promoting inappropriate research methodology for school-based contexts. These misconceptions have allowed the focus in the effectiveness research in language minority education to shift from equal educational opportunity for students to politically driven agendas.

Research Questions on Effectiveness

For 25 years, this field has been distracted from the central research questions on school effectiveness that really inform educators in their decision making. Policy makers have often chosen "Which program is better?" as the central question to be asked. But this question is not the most important one for school decision makers. Such a question is typically addressed in a short-term study. However, short-term studies, even those few that qualify as well-done experimental research, are of little or no substantive value to school-based decision-makers who vitally need information about the long-term consequences of their curricular choices. School administrators are in the unfortunate position of having to make high stakes decisions for their students now, with or without help from the research community.

A second reason for the relative lack of importance of the research question, "Which program is better?" is that **what really matters is how schools are able to assist English learners, as a group, to eventually match the achievement characteristics of native-English speakers, in all areas of the curriculum.** The U.S. Constitution's guarantees of equal opportunity, as articulated in court decisions such as *Castañeda* v. *Pickard* (1981), have come to mean that schools have an obligation to help English learners by selecting sets of instructional practices with high theoretical effectiveness, by implementing these programs to the best of their abilities and resources, and then to evaluate the outcomes of their instructional choices in the long-term. Thus, the research question of overriding importance, both legally and educationally, is "Which sets of instructional practices allow identified groups of English learners to reach eventual educational parity, across the curriculum, with the local or national group of native speakers of English, irrespective of the students' original backgrounds?"

Research Methodology in Effectiveness Studies

In addition to asking the wrong research questions, much misinformation exists regarding appropriate research methodology in the program effectiveness studies in language minority education. Reviews of research methodology issues written in politically motivated reports often focus on certain methodology issues regarding the internal validity of studies while ignoring more important methodological concerns in statistical conclusion validity and external validity. Here are some of the most common errors made in the name of "scientific" research.

Inappropriate use of random assignment. One such review from Rossell and Baker (1996) suggests that only studies in which students are randomly assigned to treatment and control groups are "methodologically acceptable." The flaw with this line of thinking is that state legislative guidelines often mandate the forms of special assistance that may be offered to language minority students, rendering impossible a laboratory-based research strategy that compares students who receive assistance to comparable students who do not. Likewise, federal guidelines based on the *Lau* v. *Nichols* (1974) decision of the U.S. Supreme Court, require that *all* English language learners receive some form of special assistance, making it unlikely that a school system could legally find a laboratory-like control group that did not receive the special assistance. At best, one might find a comparison group that received an alternative form of special assistance, but even this alternative is not easily carried out in practice.

Assuming that a comparison group can be formed, it frequently does not qualify as a control group comparable to the treatment group because school-based researchers rarely use true random assignment to determine class membership. Of those who say that they do use random assignment, most are really systematically assigning every Nth person to a group from class lists, where N is the number of groups needed. In other words, the first student on the list goes to the first program, the second student to the second program, and so on, as each program accepts the next student from the list. Since the class lists themselves are not random, but are usually ordered in some way (e.g., alphabetically), the resulting "random" assignment is not random at all, but reflects the systematic order of the original list of names. This is especially likely to result in non-comparable groups when the number of students assigned is small, as in the case of individual classrooms. Thus, what may be called random assignment is often not random in fact, if one inquires about the exact way students were assigned to treatments.

There is an additional ethical dilemma with true random assignment of students to program treatments. If the researcher knows, or even suspects, that one treatment is less effective than another, he or she faces the ethical dilemma of being forced to randomly assign students to a program alternative that is likely to produce less achievement than an alternative known to be more effective. For example, the authors, as researchers, would not randomly assign any students to ESL-pullout, taught traditionally, as a program alternative since the highest long-term average achievement scores that we have ever seen for any sizable number of students who have experienced this program, no matter how advantaged the socio-economics and other contextual variables of the schools they attended, is at the 31st NCE (or 18th national percentile) by the end of 11th grade. (See our findings later in this report.) Now that we realize how ineffective this program can be in the long term, we recommend that schools move away from this alternative completely. Certainly, we recommend that English learners not be assigned to it, randomly or otherwise, given this program's long-term lack of potential for helping them achieve eventual parity with native-English speakers.

Even a study that does succeed in establishing initially comparable groups by some means such as random assignment typically examines only very short-term phenomena and small groups. Why? Even if it were practically and ethically possible to randomly assign large groups of students to one program or another, new language minority students continually enter and others leave the schools in very non-random ways for systematic reasons (e.g., an influx of refugees, the changing demographics of local school attendance areas). When this occurs, not only does it reduce the number of "stayers" from the previous year (the internal validity problem called experimental mortality), but it can render initially comparable groups quite non-comparable within a year or two, thus destroying the "comparable groups" standard that random assignment is designed to produce. This means that studies with randomly assigned students must be short-term studies when conducted in school-based settings. Unlike the case of large medical studies of adults, we have no way to "track" students who move away and then to test them years later, in order to maintain the comparability of our initial groups. Our position is that short-term studies, with or without random assignment or other characteristics of so-called "scientific" research, are virtually useless for decision-making purposes by those school administrators and leaders who want and need to know the long-term achievement outcomes of their curricular choices now.

Statistical conclusion validity. Additional problems with research reviews of program effectiveness in language minority education center around an overemphasis on internal validity concerns, ignoring other more important issues in research methodology in education (e.g. August & Hakuta, 1997; Rossell & Baker, 1996). A common mistake is to completely ignore most or all of the factors associated with *statistical conclusion validity*--such as the effects of sample size, level of significance, directionality of hypotheses tested, and effect size on the statistical power of the research. Yet these factors are primary determinants of the research study's practical use for decision-making. Some examples of these problems are:

- *Low statistical power.* Typically small sample sizes lead to incorrect "no-difference" conclusions when a more powerful statistical test with larger sample size would find a legitimate difference between groups studied (e.g., bilingual classrooms and English-only classrooms).
- *Failure to emphasize practical significance of results over statistical significance.* The finding of statistical significance (or the lack of it) is primarily "driven" by sample size. In fact, even minuscule differences between groups can be found statistically significant using greatly inflated sample sizes. Also, enormous real differences between groups can be ob-

scured by sample sizes that are too small. A remedy for this dilemma is to report *effect size*, a measure of the practical magnitude of the difference between groups under study. One simple measure of effect size is the difference between two group means, divided by the control group's standard deviation.

But it is often difficult to form a truly comparable control group. It is sometimes possible to construct a comparison group from matched students in similar schools. If truly comparable local control groups are not available, one can construct a comparison group from the performance of other groups such as the norm group of a nationally normed test. This is facilitated through the use of NCE scores whose characteristics are referenced to the normal distribution with a mean of 50 and a standard deviation of about 21. This national standard deviation is used instead of the control group standard deviation in computing effect size. However, very few studies involving program effectiveness for English learners, whether purporting to be scientific or not, compute effect size by any method. Many researchers feel that practical significance, as measured by effect size, is much more important than statistical significance, and certainly school-based decision-makers can benefit from it to a much greater degree.

- *Violated assumptions of statistical tests.* While there are many assumptions that can be tested for a wide variety of statistical tests, research specialists are especially wary of *analysis of covariance (ANCOVA)* as recommended by "scientific" researchers to statistically adjust test scores to artificially produce "comparable" groups when it is not possible to do so procedurally, using matching or random assignment. Why? Because these researchers almost never test ANCOVA's necessary assumptions before proceeding with the adjustments to group means, making it a very volatile and potentially dangerous tool when used without regard to its limitations.

The basic problem is that ANCOVA is easy to perform, thanks to modern statistical computer programs, but difficult to use correctly. ANCOVA, when used to artificially produce comparable groups after the fact, can indeed adjust group averages, thus statistically removing the effects of initial differences between groups on some variable (e.g., family income). However, each adjustment of group means must be preceded by several necessary steps. The most important of these is that, prior to an adjustment of the group averages, it must be shown that the relationship between the covariate and the outcome measure is the same for all groups. This is a test that determines the linearity and parallelism of the regression lines that apply to each group (Cohen & Cohen, 1975; Pedhazur, 1982). Ignoring this step can easily result in an under-adjustment or over-adjustment of the group averages, thus either removing a real difference between groups or producing a difference that is not real at all!

Another common mistake made possible by easy-to-use computer software is to employ numerically coded nominal variables (classifications such as male/female) or ordinal variables (such as test scores expressed in percentiles) as covariates along with interval outcome measures. When the computer software uses non-interval variables such as these to adjust the outcome means to those that would have occurred if all subjects had the same scores on the covariates, problems in group mean adjustment can result. These problems may be addressed using more advanced forms of ANCOVA (Cohen & Cohen, 1975) but the traditional ANCOVA as executed by the default options of most conventional statistical software will generally fail to deal with these problems satisfactorily. Unfortunately, the researcher may not notice and thus will fail to realize that all of his/her group mean adjustments (and thus conclusions based on inappropriately adjusted means) have been invalidated.

The authors have engaged in and observed educational research for more than 25 years and, during that time, have seen only a small handful of studies of Title I and Title VII-funded programs that have used ANCOVA correctly or defensibly. Many statisticians claim that ANCOVA should not be used in typical non-laboratory school settings at all, and all say that it should be used with great care only by knowledgeable, statistically sophisticated social scientists. Thus, researchers who say, "We used ANCOVA to produce comparable groups," but who did not test and meet ANCOVA's assumptions, have probably arrived at erroneous conclusions.

- *The error rate problem.* Research that performs lots of statistical tests (e.g., pre-post significance tests by each grade and/or school as is typically done in program evaluations), determining each to be significant or not at a given alpha level (e.g., .05), greatly increases the likelihood of an overall Type I error (a false finding of significant difference between groups). Although the probability may be .05 (or 5%) for each statistical test, the *overall* probability of finding spuriously significant results is much greater with increasing numbers of tests. For example, the probability of finding one or more false significant differences among two groups when independently computing 10 t-tests, each with an alpha level of .05 or 5%, is about 40%. (Kirk, 1982, p. 102). For 20 independent statistical tests, the probability of finding spurious significance is about 64%.

External validity. In addition, *external validity*--the generalizability of results beyond the sample, situation, and procedures of the study--is frequently ignored by assuming that the samples, situations, and procedures of these studies apply to education as typically practiced in classrooms. In fact, the research context frequently is quite contrived, because of interventionist attempts to improve internal validity through techniques like random assignment. Thus, because of efforts to improve internal validity, the external validity of the research is reduced in experimental research, failing to help decision-makers who wish to apply research findings to their "real-world" classrooms.

Strategies exist that can help improve external validity, but these are rarely used in research studies that emphasize only selected aspects of internal validity. The easiest strategy is simply to replicate the study in a variety of school contexts, documenting the differences among the contexts and examining the same variables in each setting. A second, more sophisticated strategy is to use resampling, or the "bootstrap," a technique that uses large numbers of randomly selected resamplings of the sample to statistically estimate the parameters--the mean and standard deviation--of the population (Simon, 1993; Gonick & Smith, 1993). In other words, this approach relies on mathematical underpinnings such as the Central Limit Theorem to allow researchers to infer the true characteristics of a population (e.g. students who received ESL-content instruction in elementary school), even though the sample may be incomplete, or not a random sample of a population, or drawn from school systems that may be unrepresentative nationally. The use of these strategies would go far to compensate for the enormous practical difficulties that are involved with the selection of a truly national random sample of language minority students that is the experimental researchers' unrealized ideal.

Other internal validity concerns. While "scientific" research may address one or more types of *internal validity* problems (e.g., differential selection) using random assignment, ANCOVA, or matching, other internal validity problems frequently are unaddressed, and remain as potential explanations for researchers' findings, in addition to the treatment effect. Some examples are:

- *Instrumentation.* Apparent achievement gain can be attributed to characteristics of the tests used rather than to the treatments.
- *The John Henry effect.* The control group performs at higher levels of achievement because they (or their teachers) feel that they are in competition with the treatment group.
- *Experimental treatment diffusion.* Members of the control group (or their teachers) begin to receive or use the curriculum materials or teaching strategies of the treatment, thus blurring the distinction between what the treatment group receives and what the control group receives. This occurs frequently when supposedly English-only instructional programs adopt some of the teaching strategies of bilingual classrooms or when teachers in bilingual programs utilize less than the specified amounts of first language instruction.

In summary, self-labeled "scientific" research on program effectiveness in language minority education may only address a handful of internal validity problems, and may deal with these in impractical or inappropriate ways. Also, such studies may virtually ignore major problems with statistical conclusion validity and external validity, a fatal flaw when such research is to be used by decision-makers in school systems. These studies may often be presented in public forums in support of one political position or another in language minority education, but we encourage school systems to consider them "pseudo-scientific," rather than scientific, unless the authors make efforts to address the issues raised in this section.

Research Reviews on Program Effectiveness in LM Education

Finally, there are a number of potential problems that are associated with reviews or summaries of typical program evaluations that compare program alternatives for possible use with English learners. In particular, there are several major problems with the use of the "vote-counting" method of summarizing the results of many studies or evaluations (e.g. Baker & de Kanter, 1981; Rossell & Baker, 1996; Zappert & Cruz, 1977). Light & Pillemer (1984) describe three major problems with this deceptively simple but frequently error-prone method that divides studies into "significant positive," "significant negative," and "non-significant" outcomes and then counts the numbers in each category to arrive at an overall summary. First, vote counting typically ignores the fact that a truly non-significant conclusion should result in a vote count that reflects only 5 percent of the studies in both positive and negative categories of significance, if the probability of Type I error is .05 for all studies. If more studies fall into these categories than expected by chance, vote counting typically ignores this. Yet large numbers of both positive and negative significant findings indicate important effects of the treatment that are operating in different directions, for reasons that require additional investigation of interactions with other variables.

Second, vote counting is not statistically powerful in the conditions which permeate most of education--that is, conditions of small sample size and small effect sizes. In other words, vote counting will fail to find significant treatments most of the time in educational research under normal conditions, a fatal flaw. Third, vote counting is based on statistical significance tests, which do not tell us about the magnitude of the effect in which we're interested. Thus, the use of the vote counting method of tallying the results of reviewed studies can combine the results of large, powerful studies (i.e., in terms of statistical power) with those from small, weak studies, in effect giving equal weight to each in drawing conclusions. This can lead to serious distortions in overall findings, especially if it happens that the small and weak studies support one point of view more than the larger, powerful studies.

A more appropriate strategy is to use a weighting system that gives more credence to the large and powerful studies. A better strategy is not to use vote counting at all but to rely instead on combined significance tests that describe the pooled (combined) significance of all of the statistical tests taken together. This strategy can greatly increase the statistical power of the overall test, allowing the true effect that underlies some or all of the individual studies to emerge. An even better strategy, with fewer potential problems than significance pooling, is to use the meta-analytic technique of average effect sizes. An excellent example of this approach is Willig's meta-analytic study of the effectiveness of program alternatives for English learners (Willig, 1985). Although, like any research, it can be criticized on some points, it is worth noting that it passed very high level peer review to be published in *Review of Educational Research*, one of the most prestigious research journals of the American Educational Research Association. Thus Willig's meta-analytic synthesis carries far more weight than any vote counting research summary. In our opinion, reviewers of research in program effectiveness for English learners should abandon vote counting completely, use combined significance testing sparingly and cautiously, and emphasize the use of effect sizes as a primary means of summarizing "the bottom line" for program evaluation findings.

School-based decision-makers should be aware of the above-listed problems of vote counting as a strategy for summarizing research, and should be aware that it offers many opportunities to "tilt" the overall conclusions of the research review by judicious selection of small, weak studies that support one's point-of-view, while avoiding the consideration of large, powerful studies that are deemed "methodologically unacceptable" because of artificial standards that may apply only in limited, short-term evaluative circumstances, if they apply at all. We recommend that school-based decision-makers avoid research summaries that use vote counting and rely instead on those research summaries that use combined significance tests or meta-analytic techniques that compute average effect size.

In summary, we draw some major conclusions. **The potential effect of a program that has long-term impact on its students will probably not be detected by a short-term study.** Thus a short-term study, even if labeled "scientific" by its proponents, has virtually no relevance to the long-term issues that define second language acquisition for school and to the decisions that teachers and administrators must make. We recommend to all school district personnel that they be very wary of studies that are cited as "scientific," but which in reality represent small groups studied over a short time, in ways that ignore statistical conclusion validity and other important factors that are commonly accepted by research specialists as the hallmarks of research that is useful for decision-making. **We hold that research purporting to be scientific and intended for use in making high stakes, real-life decisions about children in school systems should emphasize most (if not all) of the hallmarks of defensible research. Further, such high stakes research should address research questions other than, "Which program is better, with all initial extraneous variables controlled?" In particular, "Which instructional practices lead to eventual achievement parity between English learners and native-English speakers?" is a research question that can and should be operationally addressed in each school system, small or large.** We will describe how school systems can do this later in this document.

Because of the above-discussed problems, and because many educators do not fully understand education research techniques, politically heated debates in education tend to be accompanied by research information that may be adequate for reaching conclusions in ideal, laboratory-like conditions, but which is totally inadequate to the needs of teachers and administrators for decision-making in the schools. Thus, educators' decisions that rely on short-term program evaluations and inappropriate "scientific" research are largely well-intentioned, seat-of-the-pants, "educated guesses" as to what works best, taking into consideration the financial constraints, the instructional resources available, and the local political climate. However, in our research, we have attempted to overcome many of these problems and to provide useful, pragmatic research information that local educators can replicate on their own data, and can use in improving the quality of the decisions that they must make.

Analyzing Program Effectiveness in Our Study

We have approached this study from a non-interventionist point-of-view by examining the instructional reality that exists in each school district, with no changes imposed on the school district for the sake of the study. In such a research context, laboratory-based strategies such as random assignment of students to different school programs are inappropriate and often impossible or impractical to implement in school settings, except possibly in a few classrooms for a study that lasts only for a relatively short time. To address many of the concerns with the limitations of short-term program evaluation, we have taken several steps. First, we have sharpened and focused the research question of "Which program is better?" by asking the question in a more refined form: "Which characteristics of well-implemented programs result in higher long-term achievement for the most at-risk and high-need student?" **We have chosen in this study to examine the highest long-term student achievement levels that we can expect to find for instructional practices**

associated with each program type, when each program type is stable and well-implemented,and when only students with no prior exposure to English are included.

Second, in our study, we have controlled some of the variables that interfere with interpretation of research results by using "blocking," first to group students using categorical or continuous variables that are potential covariates, and then later to use these groups as another independent variable in the analysis. Essentially, all student scores that fall into the same group are considered to be matched (Tabachnick & Fidell, 1989, p. 348) and the performance of each matched group within each level of program type can be compared. Each group can then be followed separately and its performance on the outcome variables (typically test scores) can be investigated separately from that of other groups of similar students. Interactions between the new independent variable represented by the blocked groups and other independent variables (e.g. type of program) can be investigated.

This strategy offers a practical and feasible means of examining comparable groups of students over the educational long term. Its advantages are that it is much more practical for large groups and long-term investigation than random assignment and that it works without the often violated and burdensome assumptions of ANCOVA. In addition, its effectiveness can approach that of ANCOVA, as the number of blocks increases beyond two (Cook & Campbell, 1979, p. 180). If the ANCOVA assumptions of linear and homogeneous regressions are not met, and this is common, it is superior to ANCOVA. In summary, this strategy is practical and pragmatic for school settings more often than ANCOVA and far more often that random assignment.

However, when the assumptions of ANCOVA could be met, we used ANCOVA as a supplement to blocking, in order to take advantage of the benefits of both techniques in situations where each works best. In some of our analyses, we used an expanded, generalized form of ANCOVA called analysis of partial variance (Cohen & Cohen, 1975). Unlike traditional ANCOVA, this analysis strategy allows for categorical covariates (e.g., free vs. reduced-cost vs. full-price lunch) as well as groups of covariates entered as a simultaneous set in order to more fully evaluate the effects of group membership (e.g., type of instructional program received by students), the effects of covariates, and the interactions among them. By these means, we have attempted to control for extraneous variables within the limits imposed by the variables that school districts typically collect, without directly changing or intervening in the instructional practice of the school districts as might be appropriate in a more "laboratory-like" context.

Third, we have used the method of sampling restriction to help control unwanted variation and to make our analyses more precise. We have done this in several ways, but primarily by focusing our attention on school districts that are very experienced in providing special services to language minority students, in order to remove the large amounts of variability in student achievement caused by poor program implementation, whatever the type of program examined. This provides a "best case" look at each program type, including programs with and without first language instructional support for students. This approach provides information on the full potential for each program to meet the long-term needs of English learners when each program is well-implemented and taught by experienced staff. This approach also provides a framework for testing the theoretical predictions of the Prism Model, in a situation in which each program is "doing all that it can do" for English learners, in terms of the four major Prism dimensions (to be presented soon).

The strategy of sampling restriction for purposes of controlling unwanted variation (thus improving internal validity of the study) does limit the generalizability of the results (external validity) to the groups studied. In other words, our findings are generalizable only to well-implemented, stable programs from school systems similar to those in our study. This is not accidental. We intended to select a purposive sample of above-average school systems. Our research study was never meant to investigate a nationally representative sample of school systems—such a sample would contain mostly "average" school systems, and would be impossibly difficult to select and analyze. From the beginning, we were interested in the question of how English language learners with no prior exposure to the English language would fare in the long term when exposed to a variety of instructional program alternatives, *all of which were well implemented by experienced, well-trained school staff.* In performing our analyses, we have additionally restricted many of our investigations to students of low socioeconomic status (as measured by their receiving free or reduced-cost lunch), thus reducing the extraneous variation typically produced by this variable as well.

All of the school districts in our study have provided a wide range of services for language minority students since the early or middle 1970s, and over the years they have hired a large number of teachers who have special training in bilingual/ESL education. The school staff are experienced and define with some consistency their approaches to implementation of the various programs. These school districts were also chosen purposefully for our study because they have collected language minority data for many years, providing information on student background, instructional services provided, and student outcomes, and because they have large numbers of language minority students of many different linguistic and cultural heritages.

By choosing only well-implemented programs in school systems with experienced, well-trained staff, we have allowed each program type examined to "be the best that it can be" within the context of its school district. Thus, our study avoids mixing results from well-implemented and poorly-implemented programs, greatly reducing the problem of confounding program implementation effects with program effectiveness. Instead, we present a picture of the long-term potential for each program type when that program is well-implemented and is operating at or near its "best."

Fourth, we have greatly increased the statistical power of our study with very large sample sizes. We have achieved these sample sizes, even when attrition reduces the number of students we can follow over several years, by analyzing multiple cohorts of students for a given length of time (e.g., seven years) between major testings. The sample figure below illustrates eight available seven-year testing cohorts for students who entered school in Grade 1, were tested in Grade 4, and who remained in school to be tested in Grade 11.

We then analyzed multiple cohorts of different students over a shorter time period (e.g., six years), followed by successive analyses of different students in multi-year cohorts down to the four-year testing interval. In doing this, we have in effect "modeled" the typical school system, where many students present on a given day have received instruction for periods of time between one and

Years and Grades of Test Administration

	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96
Cohort 8								4		6		8			11
Cohort 7							4		6		8			11	
Cohort 6						4		6		8			11		
Cohort 5					4		6		8			11			
Cohort 4				4		6		8			11				
Cohort 3			4		6		8			11					
Cohort 2		4		6		8			11						
Cohort 1	4		6		8			11							

twelve years. Typically, the shorter-term cohorts (e.g., four years) contain more students than the longer-term cohorts since students have additional opportunities to leave the school system with each passing year.

Using this approach, we are able to "overlay" the long-term cohorts with the shorter-term cohorts and examine any changes in the achievement trends that result. If there are no significant changes in the trends, we can then continue this process with shorter-term cohorts at each stage. If significant changes occur in the data trends at a given stage, we pause and explore the data for possible factors that caused the changes.

As a final step, we have validated our findings from our five participating school systems by visiting other school systems in 26 U.S. states during the past two years, and asking those school systems who had sufficient capabilities to verify our findings that generalized across our five participating districts. Thus far, at least three large school systems have conducted their own studies and have confirmed our findings for the long-term impact on student achievement of the program types that they offer. Several more have performed more restricted versions of our study and have reported findings very much in agreement with ours. This cooperative strategy considerably increases the generalizability or external validity of our findings through replication. It also allows us to make stronger inferences about how well each program type is capable of assisting its English learners to eventually approach the levels of achievement of native-English speakers in all school subjects, not just in English.

An important feature of our study is that the school districts participating in our study have been promised anonymity. The participating school systems retain ownership of their data on students and programs, allowing the researchers to have limited rights of access for purposes of collaboratively working with the school systems' staff members to interpret the findings and make recommendations for action-oriented reform from within. Our agreement states that they may identify themselves at any time but that we, as researchers, will report results from our collaborative research only in forms that will preserve their anonymity. These school systems wish to use their data to inform their teachers, parents, administrators, and policy makers, and to engage these same groups in system-wide commitments to genuinely reform their schools by improving the educational outcomes for all of their students over the next 5-10 years. Working toward this goal, they wish to emphasize the local importance of their work for the improvement of their local schools, and have little or no interest in attracting national attention until their long-term efforts have produced tangible results.

Magnitude of Our Study

Our study achieves generalizability not by random sampling, but through the use of large numbers of students from five moderate-to-large urban and suburban school systems from all over the U.S. In addition, we have added generalizability to our findings by means of replication. Specifically, we have validated our findings by comparing our results to those of other U.S. school systems in the 26 states that we have visited in the past two years. A true national random sample of language minority students (or schools) is impractically expensive to select and test, and increasingly meaningless as the underlying characteristics of the language minority population change over time. No study has ever taken this approach and none is likely to, for the practical reasons described above.

Our study includes over 700,000 language minority student records, collected by the five participating school systems between 1982 and 1996, including 42,317 students who have attended our participating schools for four years or more. This number also includes students who began school in the mid-1970s and were first tested in 1982. Over 150 home languages are represented in the student sample, with Spanish the largest language group (63 percent, overall). The total database includes new immigrants and refugees from many countries of the world, U.S.-born arrivals of second or third generation, descendants of long-established linguistically and culturally diverse groups who have lived for several centuries in what are now the current U.S. boundaries, as well as students at all levels of English proficiency development. **This represents the largest database collected and analyzed in the field of language minority education, to date.** We purposely chose to analyze school records for such a large student sample to capture general patterns in language minority student achievement. Given the variability in background among this diverse student population, including variability in the amount of their prior formal schooling, the wide range of levels of their proficiency in English, the high level of student mobility, and the variations in school services provided for these students in U.S. schools, we have found it necessary to collect substantial amounts of data to have sufficient numbers of students with similar characteristics, in order to employ our strategies for controlling extraneous variables as we follow students across time.

From this massive database, with **each school district's data analyzed separately,** we have performed a series of cross-sectional (investigating different groups of students at one or more points across time) and longitudinal analyses (following the same students across time). Also, we have analyzed multiple cohorts of students for each of several time periods. This approach acknowledges that new language-minority and native-English-speaking students are entering the school systems with each passing month and that, on a given day, the student population is made up of students who have one, two, three, or more years of instructional experience in that school system. In each analysis, we have carefully examined separately the student groups defined by each student background variable that has been collected, so that we are not comparing "apples and oranges." For example, in one series of analyses, we have chosen to look only at low-income language minority students who began their U.S. schooling in kindergarten, had no prior formal schooling, and were just beginning development of the English language.

In addition, our data analysis approach allows us to follow the directives of robust statistical analysis, which allows for stronger inferences when interesting trends in the data converge and are replicated in the variety of "data views" afforded by our analysis approach. In other words, when an initially tentative data trend or finding is first encountered, we test it by seeing whether that same trend is evident in more than one cohort, in more than one instructional setting, and in more than one time period. Trends and findings that are robust in terms of statistical conclusion validity and external validity are replicated in a variety of data views. Analytical trends and findings that are unique to a particular set of circumstances or a particular group of students are not verified across groups or across time. The findings and conclusions presented in this report have all been confirmed across student cohorts, across time periods, and across school districts.

We arrive at robust, generalizable conclusions by running the gamut of possible research investigations, from purely cross-sectional to purely longitudinal (including blended studies that combine both types, using multi-year student cohorts) for the maximum decision-making benefit of our participating school systems. Since different data views are appropriate for the wide variety of data-based reform decisions that our school systems wish to make, we and the collaborating school personnel are able to make recommendations for differential actions by teachers at the classroom level, by administrators at the school and district levels, and for policy makers at the district-wide level by referring to the data views from among our many analyses that are most appropriate for each of these audiences. For example, this approach allows the schools to investigate how their sixth grades change over the years, as well as how the 1986 third graders are doing as high school seniors in 1996, as well as how the 1985 third graders did as seniors in 1995, and how the 1984 third graders did as seniors in 1994, including dropout information.

OUR FINDINGS: THE "HOW LONG" RESEARCH

This study emerged from prior research that we had been conducting since 1985, addressing the "how long" question. In 1991, we began the current study with four large urban and suburban school districts, and in 1994, a fifth school district joined our study. Since we had already conducted a series of studies analyzing the length of time that it takes students who have no proficiency in English to reach typical levels of academic achievement of native speakers of English, when tested on school tests given in English, we chose to begin analyses of the new data from each school district by addressing this same question. The "how long" research question can be visually conceptualized in Figure 1.

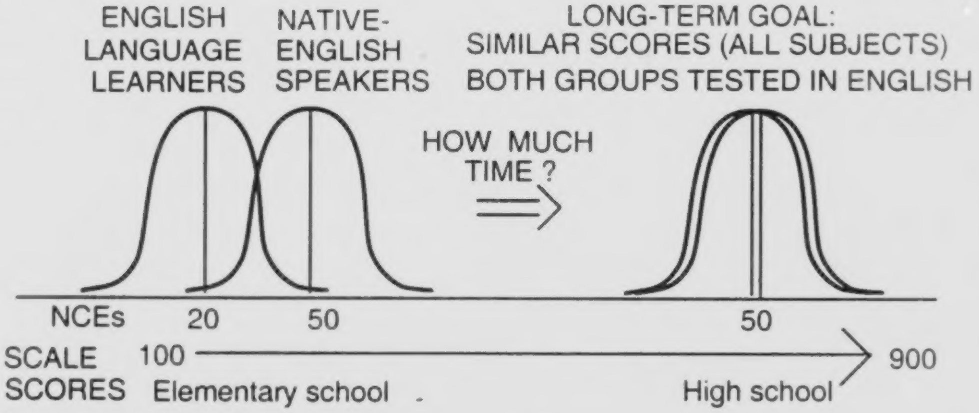
How Long: Schooling Only in L2

Our initial decision to pursue this line of research was based on Jim Cummins' (1981) study analyzing 1,210 immigrants who arrived in Canada at age 6 or younger and at that age were first

Figure 1

HOW LONG?
for students with no prior background in English
to reach typical native speaker performance on:

- norm-referenced tests
- performance assessments
- criterion-referenced measures


ENGLISH LANGUAGE LEARNERS
NATIVE-ENGLISH SPEAKERS
LONG-TERM GOAL: SIMILAR SCORES (ALL SUBJECTS) BOTH GROUPS TESTED IN ENGLISH
HOW MUCH TIME ?
NCEs
20
50
50
SCALE SCORES
100
900
Elementary school
High school

Operational definition of "equal opportunity":
The test score distributions of English learners and native English speakers, initially quite different at the beginning of their school years, should be equivalent by the end of their school years as measured by on-grade-level tests of all school subjects administered in English.

BEST COPY AVAILABLE

exposed to the English language. In this study, Cummins found that when following these students across the school years, with data broken down by age on arrival and length of residence in Canada, it took at least 5-7 years, on the average, for them to approach grade-level norms on school tests that measure cognitive-academic language development in English. Cummins (1996) distinguishes between conversational (context-embedded) language and academic (context-reduced, cognitively demanding) language, stating that a significant level of fluency in conversational second language (L2) can be achieved in 2-3 years; whereas academic L2 requires 5-7 years or more to develop to the level of a native speaker.

Since most school administrators are extremely skeptical that 5-7 years are needed for the typical immigrant student to become proficient in academic English, with many policy makers insisting that there must be a way to speed up the process, we decided to pursue this research question for several years with varied school databases in the United States. Our initial studies, first reported in Collier (1987) and Collier & Thomas (1989), took place in a large, relatively affluent, suburban school district with a highly regarded ESL program, and typical ESL class size of 6-12 students. The student samples consisted of 1,548 and 2,014 immigrant students just beginning their acquisition of English, 65 percent of whom were of Asian descent and 20 percent of Hispanic descent, the rest representing 75 languages from around the world. These students received 1-3 hours per day of ESL instructional support, attending mainstream (grade-level) classes the remainder of the school day, and were generally exited from ESL within the first two years of their arrival in the U.S.

We limited our analyses to only those newly arriving immigrant students who were assessed when they arrived in this country as being at or above grade level in their home country schooling in native language, since we expected this "advantaged" on-grade-level group to achieve academically in their second language in the shortest time possible. It was quite a surprise to find a similar 5-7 year pattern to that which Cummins found, for certain groups of students. **We found that students who arrived between ages 8 and 11, who had received at least 2-5 years of schooling taught through their primary language (L1) in their home country, were the lucky ones who took only 5-7 years. Those who arrived before age 8 required 7-10 years or more!** These children arriving during the early childhood years (before age 8) had the same background characteristics as the 8-11-year-old arrivals. **The only difference between the two groups was that the younger children had received little or no formal schooling in their first language (L1),** and this factor appeared to be a significant predictor in these first studies.

L1 schooling has now been confirmed as a key variable in our succeeding studies on the "how long" question as well as in many other researchers' work (e.g. Baker, 1993; Cummins, 1991, 1996; Díaz & Klingler, 1991; Freeman & Freeman, 1992; García, 1993, 1994; Genesee, 1987, 1994; Hakuta, 1986; Lessow-Hurley, 1990; Lindholm, 1991; McLaughlin, 1992; Pérez & Torres-Guzmán, 1996; Snow, 1990; Tinajero & Ada, 1993; Wong Fillmore & Valadez, 1986). One more age group in our initial studies, those arriving after age 12 with good formal schooling in L1, were making steady gains with each year of school, but by the end of high school, they had run out of time to catch up academically to the native-English speakers, who were constantly pulling ahead. Allowed to continue in college, though, their pattern during high school of making more gains than the native-English speaker with each year of schooling would predict that they would close the gap sometime during their undergraduate schooling. Students of all ages reached grade-level achievement in mathematics and language arts (measuring easily taught discrete points in the English language) in a shorter period of time, but required many years to reach grade level in reading, science, and social studies in English.

The measures that we use to analyze student achievement are standardized, on-grade-level, norm-referenced and criterion-referenced tests and performance assessments given in English across the curriculum--reading, language arts, mathematics, science, and social studies--the ultimate measures of attainment for eventual competition with native-English speakers on the standardized tests required for admission to a four-year university. **These tests are inappropriate measures in the first 2-3 years of English language learners' schooling in L2,** because when tested in English, the tests underestimate what these students actually know and can demonstrate when tested in L1. But eventually, **after several years of L2 schooling, these school tests in English across the curriculum become more appropriate measures to examine. These tests help parents and school administrators to know whether their children will eventually gain access to the same educational opportunities that native-English speakers have, by achieving educational parity with native-English speakers while in school.**

The insights gained from our initial studies led us to pursue the question with additional databases as well as research syntheses on other researchers' work on the "how long" question (Collier, 1987, 1988, 1989, 1992, 1995b, 1995c; Collier & Thomas, 1989; Thomas, 1992; Thomas & Collier, 1996). In all of our data analyses, as well as other researchers' work, we have continued to find the same general pattern when English language learners (ELLs) are schooled all in English and tested in English. When schooled all in English in the U.S., the shortest period of time for typical ELLs to match the achievement of typical native-English speakers is five years, among the most advantaged immigrant students who have had at least 2-3 years of on-grade-level schooling in their primary language in their home country before they arrive in the U.S. However, many ELLs schooled all in English rarely reach grade-level achievement, as measured by typical native-English speaker performance. Furthermore, **we have found that students being schooled all in English initially make dramatic gains in the early grades, whatever type of program students receive, and this misleads teachers and administrators into assuming that the students are going to continue to do extremely well.** Students are then exited from special services and it is rare for school districts to continue to monitor the ELLs' progress once they are in the mainstream, as the school work gets more cognitively complex with each succeeding grade level. Since schools usually do not monitor the progress of these students in the mainstream, the schools do not detect the fact that these students typically fall behind the typical achievement levels of native-English-speakers (defined as the 50th percentile or normal curve equivalent [NCE]) by 1-4 NCEs each year, resulting in a very significant, cumulative achievement gap of 15-26 NCEs by the end of their school years. (See Appendix A for an explanation of NCEs, their relationship to percentiles, and our rationale for their use.)

What we have found, after initial dramatic gains among most ELLs in Grades K-3, regardless of program type, is that **as these students being schooled all in English (L2) move into cognitively demanding work of increasing complexity, especially in the middle and high school years, their rate of progress becomes less than that of native-English speakers, and thus their performance, measured relative to native-English speaker performance in NCEs, goes down.** As a group, the typical performance of ELLs schooled exclusively in English reaches its maximum at a level substantially below the 50th percentile or NCE, the typical performance of the native-English speaker. **It is important to understand that typical students in all program groups achieve significant gains each year. But when comparing groups, English language learners who have received all their schooling exclusively through L2 might achieve 6-8 months' gain each school year as they reach the middle and high school years, relative to the 10-month gain of typical native-English speakers.** Thus, an achievement gap with native-English speakers that was partially closed in elementary school becomes wider with each passing year, as typical native-English speakers inexorably advance by making 10 months' gain in 10 months' time, to maintain their average score at the 50th NCE across the years. In these analyses we have examined performance on standardized norm-referenced and criterion-referenced tests, local school district measures, and state performance assessments, and whatever the measure, we find the same general pattern, when students are tested in the cognitively complex subjects as they leave the early childhood years.

How Long: Schooling in Both L1 & L2

After continuing to hear the insistent voice of policy makers to find a way to "speed up" or accelerate the process, in our next studies, we began to examine the progress of students in bilingual programs. Could the process of bilingual schooling speed up the acquisition of academic L2 and academic achievement in general?

What we found again was quite a surprise. We limited our analyses to students attending well-implemented bilingual classes taught by experienced bilingual teachers, and used as a measure of consistency the students' level of academic achievement in their first language. Those students on grade level in L1 (i.e., tested in math, science, social studies, and language arts in L1) reached on-grade-level performance in English (L2) in all subject areas in 4-7 years.

At first these data analyses appeared to present a rather bleak picture--that it takes a long, long time whatever the program--until we examined the long-term picture for Grades K-12 with additional data from our current study with five large, experienced school districts. What we have found is that following these students throughout their schooling, **the bilingually schooled students are able to sustain the gains in L2, and in some cases, to achieve even higher than typical native-English-speaker performance as they move through the secondary years of school.** In other words, once bilingually schooled students "get there" (where "there" is parity with comparable native-English speakers of similar age on the school tests in English), they stay there, achieving on or above grade level in L2. In contrast, the students who have been schooled all in English (L2), tend to go back down in achievement (i.e., lose ground relative to native speakers of English) as they reach the upper grades of school. The students schooled only in L2 do not sustain the gains they made during the elementary school years, when compared to typical native-English speaker gains across the years.

Figure 2 illustrates the range of student performance on English reading, demonstrating the dramatic difference between the performance of students who receive grade-level academic work in L1 and those who do not receive L1 instructional support after their arrival in the U.S.

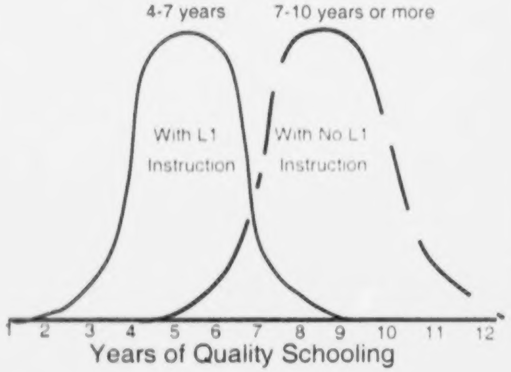
As can be seen in the figure, a few students achieve above and below the group patterns, but typical students schooled bilingually who reach the 50th NCE (or on-grade-level performance in English) received no L1 instructional support leave school without high school completion; whereas many more students who were schooled bilingually graduate from high school. (We will present the dropout data in future reports.)

Figure 2

How Long?

(to reach 50th NCE on
English Reading subtest in L2
with no prior English exposure)

with L1 instruction: 4-7 years
with no L1: 7-10 years or more



©Copyright Wayne P. Thomas, 1997

Summary of "How Long" Findings for English Language Learners

So it takes typical bilingually schooled students, who are achieving on grade level in L1, from 4-7 years to make it to the 50th NCE in L2. It takes typical "advantaged" immigrants with 2-5 years of on-grade-level home country schooling in L1 from 5-7 years to reach the 50th NCE in L2, when schooled all in L2 in the U.S. It takes the typical young immigrant schooled all in L2 in the U.S. 7-10 years or more to reach the 50th NCE, and the majority of these students do not ever make it to the 50th NCE, unless they receive support for L1 academic and cognitive development at home.

How Long: Bilingual Schooling for Native-English Speakers

Next we examined native-English speakers whose parents chose to have their children placed in a two-way bilingual class. These students include those with many advantages. For example, their first language, English, is not threatened in any way. English is the status and power language of the U.S., as well as of the world. We have examined English-speaking Euro-American children of middle and lower income homes, as well as African-American children of middle and lower income homes who have chosen to attend bilingual classes. The middle-income children often

BEST COPY AVAILABLE

have parents cheering them on, providing L1 cognitive and academic support at home. How long does it take these "advantaged" English speakers? Four to seven years is the minimum time frame for these students to reach the point where they can show off what they know on the school tests in their second language, at the level of a native speaker of that language. These middle-income students achieve on or above grade level in English (L1) with each year of school, but it still takes until at least fourth or fifth grade for the typical students in this group to make it to the 50th NCE on school tests in L2. Once they get there, they stay there and can demonstrate what they know in either L1 or L2, as long as L2 grade-level academic work continues to be provided. In other researchers' studies, in the U.S. as well as other countries, similar results have been found around the world when following bilingually schooled students long-term (Collier, 1992; Lindholm, 1990; Lindholm & Aclan, 1991). Typical low-income native-English speakers, including low-income African-American students, also generally reach and stay at the 50th NCE in L1 within 4-7 years of bilingual schooling, and can achieve at the 50th NCE in L2 if schooling is continued in L2.

How Long: Influence of Student Background Variables

Proficiency in L1 & L2

In each of these "how long" studies, we have examined groups of students separately, grouping by student background variables that the bilingual/ESL staff in each school system have identified as having potential influence on student achievement. One variable that we have found extremely important to examine in separate groupings is students' level of proficiency in the language of instruction. We have assessed the influence of this variable, proficiency in L1 and L2, by the age and grade level of the student, by the language proficiency measures used by each school system, by the level of L1 and L2 instruction in which students are placed in each school year, and by the number of months/years of exposure to the language of instruction. To avoid "mixing apples and oranges," we have accounted for this variable by analyzing similar groups of students of the same age who start L2 proficiency development at the same point in time. For example, we might follow the progress of a group of Spanish-speaking ESL beginners in first grade who receive one type of instructional support in the initial years, following them over time for as many years as they remain in the same school system. New arrivals who speak the same L1 who arrive in *second* grade with no proficiency in English and receive one type of instructional support are a second group that we follow, and so on. We have found that the number of years of exposure to the English language is a strong predictor of ELLs' long-term academic achievement, so it is very important to account for this variable. **All groups, whatever their circumstances, demonstrate growth in development of the English language with each additional year of exposure to the language.**

Age

Age of the student is a parallel variable with language proficiency level. The language system that a five-year-old uses is very different from that of a thirteen-year-old. An eighteen-year-old has developed a fairly mature system of language, but even the young adult will continue to expand vocabulary, pragmatics, discourse, and writing competence throughout life. So in our analyses, we always examine separately each possible combination of age and language proficiency.

Student's First Language (L1)

Another student background variable that many teachers assume has an influence is the **student's L1.** We have found that the particular L1 that a student speaks is not a powerful variable in long-term academic achievement. In other words, we have found that Spanish speakers make the same rate of progress in L2 as do speakers of Arabic or Mandarin Chinese or Amharic or Korean or Russian or Vietnamese. But there is a relationship between L1 and L2. The true predictor is not *which* first language the student speaks but how much cognitive and academic development in L1 the student has experienced. **The deeper a student's level of L1 cognitive and academic development (which includes L1 proficiency development), the faster students will progress in L2.** This generalization can be verified by numerous studies examining age and its relationship to L2 proficiency development. Many researchers have found that older students are more efficient than younger children in the acquisition of L2 (see Collier, 1988, 1989 for syntheses of studies on the age issue). In our current study, we have found that formal schooling in L1 is the true predictor, not which particular first language the student happens to speak. ESL teachers who describe differences in rate of ESL acquisition among their different language groups are experiencing only short-term differences. In the long term, we have found that differences between language groups disappear when comparable groups, with same levels of L1 schooling, are compared to each other.

Socioeconomic Status

Does socioeconomic status (SES) make a difference? Overall, SES is a powerful predictor of school achievement in many research studies in education. But in our study we are finding that this is a difficult variable to measure. All of our school districts collect this variable in school records by identifying those students who qualify for free or reduced lunch. This provides an indirect and gross measure of family income. We have found that a majority of language minority students in our data base have qualified for free or reduced lunch, approximately 57 percent of our sample. But these students do not always experience life similar to that of the average family in poverty in the U.S. This category of students is highly variable. For example, some are recently arrived immigrants who have begun life over again in this country, having emigrated from an economically depressed or war-torn area of the world. But some of these new arrivals were well-educated, middle income families in their country of origin who experience temporary income reduction after immigrating. While it might take these parents ten years in the U.S. to attain the professional credentials to continue work at their former level of income, they have the aspirations and education of the middle class, and they have given their children the L1 cognitive and academic support needed for the students to be on grade level when they arrive. Even when we look only at the records of low-SES students (using the criterion of free/reduced lunch), we find that this variable is confounded with other variables, such as family aspirations, previous SES in home country, and amount of parents' formal schooling.

In addition, when we attempt to control the effects of SES by investigating only low-SES students, we find that student-level differences between school programs have far more explanatory power for predicting student achievement than family background differences among students.

For these reasons, we have found that the generalized SES variable is less useful than we anticipated as an explanatory student background variable. Again, **the more powerful variable is more specific and educationally relevant—the amount of formal schooling in L1 that the student has experienced.** We also have some survey data from two of our school districts that indicates that the amount of formal schooling parents have completed can be a significant predictor of their childrens' academic success in the U.S. Both of these variables—student schooling in L1 and formal schooling of parents— may be moderately correlated with family income and so one might initially attribute their effects to the more general variable of student SES. However, we believe that it is more appropriate to measure and evaluate these two variables as more direct influences on student achievement than SES. **Formal schooling is the true predictor.** In contrast to the powerful predictors of student schooling in L1 and to parental education levels, **parents' level of proficiency in English is not an important predictor of student achievement in English.**

Finally, we have found that, within the group of low-SES students (who represent the majority in our database), school program is a very powerful predictor of school achievement. This is also true when we investigate the total sample of students of all SES levels. Thus, it appears that school program can "explain" or "capture" as much (and usually more) variance in student achievement as is explained by SES. In effect, the differences in school programs are more powerful at explaining student achievement than SES. We conclude that the selection of the most effective programs for English learners can provide for long-term school achievement for even the students of lowest SES backgrounds. In fact, our databases contain several hundred student records from several two-way bilingual schools in very economically depressed neighborhoods where the school's programs and teachers have successfully assisted many low-SES students to dramatically outscore their more economically privileged peers and even to outscore typical advantaged native speakers of English in the same school systems. A school's well-implemented bilingual program for English learners can indeed overcome the effects of low SES on long-term student achievement.

Formal Schooling in L1

In our review of other researchers' work on long-term academic achievement in L2 that we conducted at the beginning of this current study, we created the following generalization to summarize other researchers' findings: "The greater the amount of L1 instructional support for language minority students, combined with balanced L2 support, the higher they are able to achieve academically in L2 in each succeeding academic year, in comparison to matched groups being schooled monolingually in L2" (Collier, 1992). After analyzing over 700,000 student records in our current five school district sites, to answer the "how long" question, we find that this generalization still holds. **Of all the student background variables, the most powerful predictor of academic success in L2 is formal schooling in L1. This is true whether L1 schooling is received only in home country or in both home country and the U.S.**

UNDERSTANDING OUR "HOW LONG" FINDINGS: THE PRISM MODEL

Why does it take so long? Why do so many students schooled only in L2 rarely reach the 50th percentile on norm-referenced tests? Why do so few ELLs ever reach the typical performance of the native-English speaker on performance assessments or criterion-referenced tests, even when they are given intensive course work all in English? Why does it take typical bilingually schooled students a minimum of 4 years, and as long as 7 years, to "demonstrate what they know" in their L2, at typical performance levels of native speakers?

When we first began reporting on this data and interpreting the results, we discussed second language proficiency development as the main reason for students' low performance. We said that it takes many years to develop academic English. **Now we interpret our findings differently. Second language acquisition is only one of many processes taking place.** Does it take 4-10 years or more to acquire a second language for schooling purposes? Clearly language acquisition is a complex, developmental process. But **the main reason that it takes so long for ELLs to reach grade-level performance on tests in English is that native-English speakers are not standing still waiting for ELLs to catch up with them** (Thomas, 1992). **Native-English speakers are developing cognitively and academically with every year of school, as well as continuing their acquisition of L1 in a learning environment that is favorable for instruction in English. School tests reflect that ongoing linguistic, cognitive, and academic growth that occurs in an "English-friendly" learning environment.**

The Instructional Situation for the Native-English Speaker

Examining what happens developmentally to the native-English speaker in school provides insights into the complex developmental processes also occurring for the non-native speaker of school age. It also helps us understand the results from the tests that we use to measure progress in school. All children experience natural, complex developmental processes that are ongoing throughout the school years. Two major developmental processes that occur at the subconscious level are linguistic and cognitive development, and these ongoing processes can be stimulated by consciously planned activities with teachers, parents, siblings, and friends. Language and cognitive development go hand in hand. Language is the vehicle for communicating cognitively. In school, we develop students' cognitive growth through academic work across the curriculum in science, social studies, mathematics, language arts, and the fine arts. At home, parents naturally stimulate children's cognitive growth through daily, interactive problem-solving, family activities, and household responsibilities. All of this growth at home and school, conscious and subconscious, is reflected in the school tests, especially when long-term student progress is followed, with different tests for each age group or grade level. Teachers' tests change from week to week to reflect this expected growth. School district, state, and nationally normed tests change from year to year to reflect this expected growth.

Another perspective to provide insight on what the school tests measure is to understand the continuous L1 developmental process that is ongoing throughout the school years for native-English speakers. Often it is assumed that the five-year-old native-English speaker entering school is fully proficient in the English language. This child is amazingly adept in using a complex oral language system, developed cognitively to the level of a five-year-old. But even for the most gifted five-year-old, much more than half of the English language remains to be acquired during the school years. Children from ages 6 to 12 continue to acquire (without being formally taught) subtleties in the phonological system, massive amounts of vocabulary, semantics (meaning), syntax (grammar), formal discourse patterns (stretches of language beyond a single sentence), and complex aspects of pragmatics (how language is used in a given context) in the oral system of the English language (Berko Gleason, 1993). Then there is the *written* system of English to be mastered across all of these same domains during the school years! Even an adolescent entering college must continue to acquire enormous amounts of vocabulary in every discipline of study and ongoing development of complex writing skills.

Once again, the school tests reflect this expected English language growth with every year of school. ELLs taking an English as a second language proficiency test are being tested on a static measure, an important indicator of growth in each of the domains of the English language. But in the meantime, native-English speakers are acquiring English too, developmentally expanding their language system with each year of school. The school tests reflect this age-appropriate growth, but the static language proficiency tests do not. ELLs are competing with a moving target when they take the school tests in English language arts and English reading. In fact, the average score on these tests is defined by the native-English speaker who makes "one year's progress in one year's time" and thus sets the standard for progress for the English learner.

This L1 language development is deeply interrelated with cognitive development. Children who stop cognitive development in L1 before they have reached the final Piagetian stage of formal operations (somewhere around puberty), run the risk of suffering negative consequences, as measured by school tests. Many studies, including this one, indicate that if students do not reach a certain threshold in their first language, they may experience cognitive difficulties in the second language (Collier, 1987; Collier & Thomas, 1989; Cummins, 1976, 1981, 1991; Dulay & Burt, 1980; Duncan & De Avila, 1979; Skutnabb-Kangas, 1981; Thomas & Collier, 1996). Furthermore, developing cognitively and linguistically in L1 at least throughout the elementary school years provides a knowledge base that transfers from L1 to L2. When schooling is provided in both L1 and L2, both languages are the vehicle for strong cognitive and academic development. Linguistically, deep structure in L1 transfers to L2. Literacy skills transfer from L1 to L2, even when L1 is a non-non-Roman-alphabet language and L2 is English (Chu, 1981; Cummins, 1991; Thonis, 1994). Cognitive processes developed in L1 transfer to L2 (Bialystok, 1991).

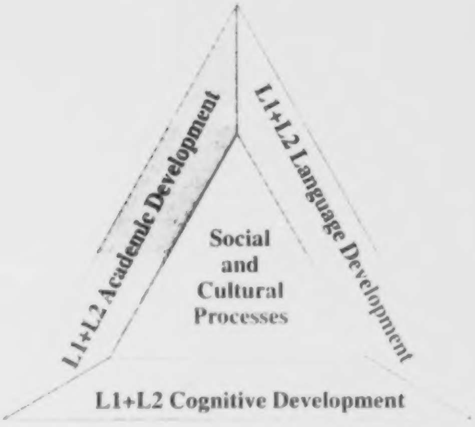
Thus, the simplistic notion—that all we need to do is to teach language minority students the English language—does not address the needs of the school-age child. Furthermore, when we teach only the English language, we are literally slowing down a child's cognitive and academic growth, and that child may never catch up to the constantly advancing native-English speaker!

The Prism Model: Language Acquisition for School

To help policy makers understand the complex process of second language acquisition within a school context, we have developed a conceptual model that has emerged from our research findings, as well as other researchers' work. (For research syntheses, see Collier, 1995a, 1995b, 1995c.) The model has four major components that"drive" language acquisition for school: sociocultural, linguistic, academic, and cognitive processes. To understand the interrelationships among these four components, Figure 3 symbolizes the developmental process that occurs during the school years for the bilingual child. While this figure looks simple on paper, it is important to imagine that this is a multifaceted prism with many dimensions. The four major components—sociocultural, linguistic, academic, and cognitive processes—are interdependent and complex.

Figure 3

Language Acquisition for School


L1+L2 Academic Development
L1+L2 Language Development
Social and Cultural Processes
L1+L2 Cognitive Development

The Prism Model

Sociocultural Processes

At the heart of the figure is the individual student going through the process of acquiring a second language in school. Central to that student's acquisition of language are all of the surrounding social and cultural processes occurring through everyday life within the student's past, present, and future, in all contexts--home, school, community, and the broader society. For example, sociocultural processes at work in second language acquisition may include individual student variables such as self-esteem or anxiety or other affective factors. At school the instructional environment in a classroom or administrative program structure may create social and psychological distance between groups. Community or regional social patterns such as prejudice and discrimination expressed towards groups or individuals in personal and professional contexts can influence students' achievement in school, as well as societal patterns such as subordinate status of a minority group or acculturation vs. assimilation forces at work. These factors can strongly influence the student's response to the new language, affecting the process positively only when the student is in a socioculturally supportive environment.

BEST COPY AVAILABLE

Language Development

Linguistic processes, a second component of the model, consist of the subconscious aspects of language development (an innate ability all humans possess for acquisition of oral language), as well as the metalinguistic, conscious, formal teaching of language in school, *and* acquisition of the written system of language. This includes the acquisition of the oral and written systems of the student's *first and second languages* across all language domains, such as phonology, vocabulary, morphology, syntax, semantics, pragmatics, discourse, and paralinguistics (nonverbal and other extralinguistic features). To assure cognitive and academic success in *second* language, a student's *first* language system, oral *and* written, must be developed to a high cognitive level at least through the English to a level comparable to their native-English-speaking peers.

Academic Development

A third component of the model, academic development, includes all school work in language arts, mathematics, the sciences, and social studies for each grade level, Grades K-12 and beyond. With each succeeding grade, academic work dramatically expands the vocabulary, sociolinguistic, and discourse dimensions of language to higher cognitive levels. Academic knowledge and conceptual development transfer from the first language to the second language. Thus, it is most efficient to develop academic work through students' first language, while teaching the second language during other periods of the school day through meaningful academic content. In earlier decades in the U.S., we emphasized teaching second language as the first step, and postponed the teaching of academics. Research has shown us that postponing or interrupting academic development is likely to promote academic failure in the long-term. In an information-driven society that demands more knowledge processing with each succeeding year, students cannot afford the lost time in on-grade-level academic work during the period while they are learning English to a level comparable to their native-English-speaking peers.

Cognitive Development

The fourth component of this model, the cognitive dimension, is a natural, subconscious process that occurs developmentally from birth to the end of schooling and beyond. An infant initially builds thought processes through interacting with loved ones in the language of the home. This is a knowledge base, an important stepping stone to build on as cognitive development continues. **It is extremely important that cognitive development continue through a child's first language at least through the elementary school years.** Extensive research has demonstrated that children who reach full cognitive development in two languages (generally reaching the threshold in L1 by around age 11-12) enjoy cognitive advantages over monolinguals. Cognitive development was mostly neglected by second language educators in the U.S. until the past decade. In language teaching, we simplified, structured, and sequenced language curricula during the 1970s, and when we added academic content into our language lessons in the 1980s, we watered down academics into cognitively simple tasks, often under the label of "basic skills." We also too often neglected the crucial role of cognitive development in the first language. Now we know from our growing research base that we must address linguistic, cognitive, and academic development equally, through both first and second languages, if we are to assure students' academic success in the second language. This is especially necessary if English learners are ever to reach full parity in all curricular areas with native-English speakers.

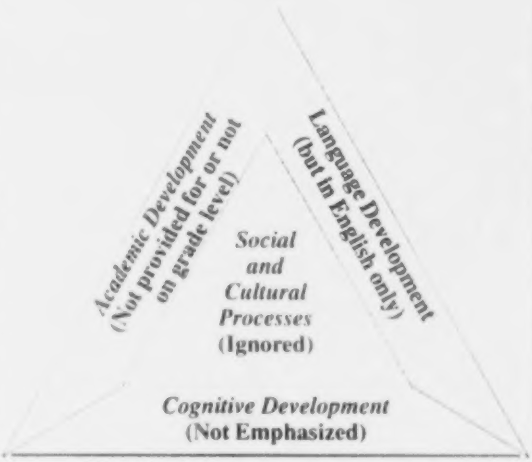
Interdependence of the Four Components

All of these four components--sociocultural, academic, cognitive, and linguistic--are interdependent. If one is developed to the neglect of another, this may be detrimental to a student's overall growth and future success. The academic, cognitive, and linguistic components must be viewed as developmental. For the child, adolescent, and young adult still going through the process of formal schooling, development of any one of these three components depends critically on simultaneous development of the other two, through both first and second languages. Sociocultural processes strongly influence, in both positive and negative ways, students' access to cognitive, academic, and language development. It is crucial that educators provide a socioculturally supportive school environment that allows natural language, academic, and cognitive development to flourish in both L1 and L2.

The Instructional Situation for the English Language Learner in an English-only Program

Figure 4

Second Language Acquisition for School : Common View of Policy Makers



The English-Only Perspective: Learn English First!

© Copyright Wayne P. Thomas & Virginia P. Collier, 1997

Using all the components of the Prism Model, we can apply this research knowledge base to the varying school programs provided for ELLs in the United States. The common view of many U.S. education policy makers is portrayed in Figure 4--that students must learn English first.

From a common-sense perspective, it would seem obvious that the first step anyone should take when entering a new country is to learn the language of that country. This is indeed a wise decision for a cognitively mature adult who has already mastered the requisite academic material to an adult level in first language. The adult immigrant who has been formally schooled has completed development in two of the three Prism dimensions--cognitive and academic development--and lacks only a portion of the linguistic dimension--acquisition of the second language--having already acquired L1 to adult level of proficiency.

But the school-age child is in a very different situation. **Developmental processes must continue nonstop all through the school years in order for a child to reach the cognitive maturity of an adult.** Academic development must continue nonstop through the school years for full adult mastery of the academic curriculum to occur. English is only one part of the learning process. When learning English is the first goal, during the period that this goal is the priority, the full Prism Model of language acquisition for school is reduced to mainly one dimension, development of one language (L2), and half of that dimension is missing--continuing development of L1. This has unhappy consequences for the student in three out of four of the Prism Model's dimensions.

First, meaningful academic development is not provided for in the initial years, because the highest priority is learning English rather than academic content. In succeeding years academic development is often not on grade level, because students studying all in L2 have missed at least two years of academic work while acquiring a basic knowledge of L2. Second, cognitive development is not emphasized in second language and is not provided for in first language at school. Students enter school having completed six years of cognitive development in their first language. These students must continue to develop cognitively at the same rate as of do native-English-speaking students in their native language. Switching a student's language instruction to all-English causes a cognitive slowdown for English learners that can last for several years. During this period, the English speakers continue to develop cognitively at normal rates but the English learners fall behind in cognitive development and may never catch up to the constantly advancing English speakers. Third, in an English-only environment, sociocultural processes may be largely ignored, or less well provided for, and thus when students feel that they are not in a supportive environment, less learning takes place.

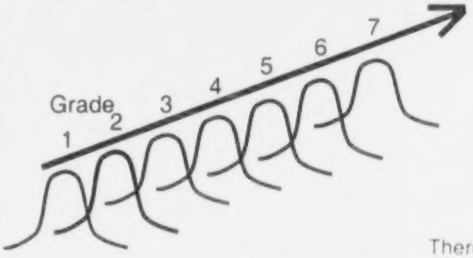
Now contrast this with the situation of the native-English speaker. For most native-English speakers, all four dimensions of the Prism Model are in place in L1, including schooling in a socioculturally supportive environment. From kindergarten on, native-English speakers are instructed in L1. Even for those who choose to participate in a bilingual class, they do not fall behind in other school subjects while learning another language during the school years. Typical native speakers of English make ten months' progress in school achievement for each 10-month school year. This performance defines the 50th percentile or NCE on standardized norm-referenced tests and the average score on criterion-referenced tests as the students progress from grade to grade. Likewise, on a state or school district performance assessment, the standards developed for each grade level are also based on typical performance of groups of native-English speakers on these tests. These tests measure continuous linguistic, cognitive, and academic growth, and the tests change weekly, monthly, and yearly to reflect that growth.

It is on these school tests that we unrealistically expect ELLs to be able to demonstrate miraculous growth. Policy makers assume that non-English-proficient students should somehow be able to leap from the first percentile or NCE to the 50th (as compared to native speakers of English) in 1-2 years. During this period, the native speakers continue to make ten months' progress in ten months' time. Yet if English language learners are being taught only in English, a language they do not yet understand, they need at least 2-3 years to reach a high enough level of proficiency in L2 to attempt to keep up with the pace of the native-English speaker in school. For example, one group of non-English-proficient students might study English intensively, and by the end of their first two years, they make an enormous leap from the first to the 20th NCE when they first take a standardized test in English reading, English language arts, and mathematics. To score at the level of the typical native-English speaker (50th percentile or NCE) in all school subjects, these English language learners must then continue to make *more* than one year's progress in one year's time, and do so for several consecutive years, to ever close the initial gap of 25-30 NCEs. For ELLs, progress at the typical rate of native-English speakers means *maintaining* the initial large gap, not closing it, as the native-English speakers continue to make additional progress in all Prism dimensions with each passing year. If ELLs make less than typical native-English speaker progress (e.g., ELLs might make 6 months' progress in one 10-month school year while typical native speakers make 10 months' progress), the initial large achievement gap will widen even further. Figure 5 visually illustrates this point.

Figure 5

AN IMPORTANT UNDERSTANDING

Typical English Speakers (50th percentile or NCE) make one year of achievement gain during each school year (10 months gain in a 10 month school year) FOR EACH YEAR OF SCHOOL



Therefore:
English language learners must typically gain MORE THAN ONE YEAR'S ACHIEVEMENT (e.g., 15 months gain) in each of SEVERAL CONSECUTIVE SCHOOL YEARS to ever close the initial 30 NCE achievement gap with English speakers WHEN TESTED IN ENGLISH (L2).

© Copyright Wayne P. Thomas, 1997

To illustrate further, if a group of English language learners experiences an initial three-year gap in achievement assessed in English (math, science, social studies, language arts, reading), they must make an average of about one-and-a-half years' progress in the next six consecutive years (for a total of nine years' progress in six years--a 30-NCE gain, from the 20th to the 50th NCE) to reach the same long-term performance level that a typical native-English speaker reaches by making one year's progress in one year's time for each of six years (for a total of six years' progress in six years--a zero-NCE gain, staying at the 50th NCE). This is a difficult task indeed, even for an English language learner who has received excellent formal schooling before entering U.S. schools and whose achievement is on grade level for his/her age when tested in his/her native language. Still more daunting is the task of the English language learner whose schooling has been interrupted by social or economic upheaval or warfare. Learning English while keeping up with native speakers' progress in other school subjects *and* while making up the material lost to interrupted or non-existent schooling in the student's native country is a truly formidable undertaking.

It is for these reasons that on-grade-level bilingual schooling is essential to these students' long-term academic success. While the student is making the gains needed with each succeeding year to close the gap in performance on the tests in English, that bilingual student is not getting behind in cognitive and academic development. Once the bilingual students' average achievement reaches the 50th NCE (the average achievement level of native-English speakers) on the school tests in English, the cognitive and academic work in L1 has kept these students on grade level and they sustain grade-level performance in English even as the academic work gets increasingly complex with each succeeding year in middle and high school.

OUR FINDINGS: SCHOOL EFFECTIVENESS

The second major research question of our study focuses on school program and instructional variables and their influence on the long-term academic achievement of language minority students. We have sought to answer this question by many different data analyses that separate similar groups of students by each student background variable, and each program treatment. We then follow these student cohort groups across time for as long as the students remain in a given school system. As with the "how long" question, we have analyzed data from each school system separately. We first examined patterns in one school district, to see if any program or instructional variables appeared to have strong influence on language minority students' achievement. Then if a particular pattern emerged, we assumed it was not generalizable beyond the context of that school system, unless we found a similar pattern in a second school system. Once the same pattern appeared repeatedly in the data across more than two research sites, we started to assume some generalizability. The patterns that we are reporting here are general academic achievement patterns across all five of our research sites. These student achievement patterns are strongly influenced by the type of school program provided by the schools in our study. In fact, we found that the schools with the highest achievement levels were so effective that the effect of these programs overcame the power of student background variables such as poverty. Low-income students were able to be high achievers in the most effective programs.

Characteristics of Effective Programs

To measure school program effectiveness in each school district, we began by interviewing school staff to identify and reach a consensus on definitions of programs and their implementation. We did this through focus groups with bilingual/ESL teachers and resource staff. Through these group interviews, we uncovered differences from one school district to another in the labels given to programs, but consistency in general characteristics of differences between programs. We have chosen here to report these findings by using the names of general program labels in bilingual/ESL education. However, **we caution the field of bilingual/ESL education not to focus so much on the name or label of a given program, but instead to think about the underlying characteristics that lead to a given program's success.** Thus we shall begin this discussion of program effectiveness with program characteristics that we have found to be very effective, rather than naming specific program models as most effective. Following this discussion, we will illustrate these effective program characteristics as they appear in some common program models in bilingual/ESL education.

L1 Instruction

It is very clear from all of our findings in this study, as well as other researchers' work, that when students have the opportunity to do academic work through the medium of their first language, in the long term they are academically more successful in their second language. In this study, students who emigrated to the U.S., after having received several years of on-grade-level schooling in their home country, made greater progress than similar groups of students who emigrated at a young age and received all their schooling in English (L2) in the U.S. Students who were born in the U.S., who received 5-6 years of on-grade-level schooling in both L1 and L2 in U.S. schools (with the remaining school years all in L2), made greater progress than similar groups who received 2-3 years of schooling in both L1 and L2 in U.S. schools (with the remaining school years all in L2). Students born in the U.S., who received 2-3 years of schooling in both L1 and L2 in U.S. schools, made greater progress than similar groups who received all of their schooling in English (L2), with ESL support, in U.S. schools. Comparing all of these groups receiving support services that differed chiefly by the amount of L1 academic support, **the message from our findings is overwhelmingly clear that all language minority groups benefit enormously in the long-term from on-grade-level academic work in L1. The more children develop L1 academically and cognitively at an age-appropriate level, the more successful they will be in academic achievement in L2 by the end of their school years.**

It is important here to remember the point made in the "how long" discussion that these findings are different from the short-term findings. Most of the studies of school effectiveness in bilingual/ESL education have focused on a short-term look at Grades K-3. And many of these studies have concluded that there is little difference between programs in the early grades. We found similar patterns in our data, but as we continued to follow groups of students through the middle and high school years, we found very large, cumulative, long-term differences in student achievement that were directly attributable to the type of program services that they received during their elementary school years. We have concluded that L1 cognitive and academic development is a key predictor of academic success in L2.

It is also important to remember that this predictor is much more powerful when L1 development is thought of as **academic enrichment through L1 age-appropriate schooling.** Some forms of bilingual education in the U.S. have focused on minimal L1 support, such as L1 literacy development. While any L1 development is beneficial, for students to get the full power of this predictor, they need to be challenged academically across the curriculum through L1. They need to do cognitively complex school tasks appropriate for their age in L1. It is possible that parents can provide some of this L1 cognitive and academic support at home. We have some survey data that suggests that parents who have completed at least a high school degree do try to provide some extra L1 support. But long work hours and the necessity to have at least two income providers in every household make this parental role increasingly difficult. When schools can provide L1 cognitive and academic support, all language minority students will greatly benefit. **This predictor holds true for immigrants to the U.S., as well as for U.S.-born language minority students. L1 schooling is powerful for students who have lost their L1, for bilingual students who are very proficient in L1 and L2, and for students who are just beginning development of L2.**

L2 Instruction

The type of L2 instructional support is the key to this predictor having power. During the portion of the school day that is taught through English (L2), we have found that it is not enough just to teach the English language. More English is not necessarily better. **L2 must be used to provide students access to the full curriculum, through ESL content, or sheltered academic instruction. But ESL taught through academic content must also be provided in a socioculturally supportive environment, while challenging students to work at age-appropriate level through L2.**

Just as with L1 instructional support, we have found a hierarchy of services for the L2 instructional component that can predict long-term academic success. But this hierarchy is not parallel to L1 instruction in all aspects. The major predictor here is that the L2 component of the day should be taught through cognitively complex academic work across the curriculum, while making the material meaningful for students at their proficiency level in L2. Thus in our study, students who received L2 taught through academic content (by teachers trained in second language acquisition and the content area, who were also socioculturally supportive of students) made greater progress than students receiving ESL classes focused on the teaching of the English language and the remaining L2 portion of the day in mainstream classes. Students who received L1 academic content and L2 academic content (taught by teachers trained in second language acquisition and the content areas who were also socioculturally supportive of students) did better than students who received only L2 academic work. We will discuss time spent in the L2 mainstream in the sections that follow.

Interactive, Discovery Learning and Other Current Approaches to Teaching

We have found that across all program types, students who participate in classes that are very interactive, with discovery learning facilitated by teachers so that students work cooperatively together in a socioculturally supportive environment, do better than those attending classes taught more traditionally. Teachers in the focus groups in our study expressed excitement when staff development sessions assisted them with cooperative learning, thematic lessons, literacy development across the curriculum, process writing, performance and portfolio assessment, uses of technology, multiple intelligences, critical thinking, learning strategies, and global perspectives infused into the curriculum. Since the teachers described this as an influence on their teaching styles, in our data analyses, we attempted to measure this change. We found that students who attended classes taught by teachers who had been through intensive staff development in these current approaches to teaching made faster long-term progress than students attending more traditionally taught classes.

To measure this predictor, in our interviews with school staff, we found that in each school district, the bilingual/ESL staff could identify a specific time period when staff development on these topics was initiated, generally in the mid-to-late 1980s or early 1990s. We examined language minority student progress in each school building prior to this intensive staff development and in the years following. Student performance was enhanced and sustained as the students moved on through school. Generally, for all program models, these changes in teaching styles resulted in a cumulative 8-10 NCE gain by 11th grade, for most students. Thus another powerful predictor of long-term student success is change to more current approaches to teaching, fostering active rather than passive classrooms.

Sociocultural Support

Sociocultural support is a difficult predictor to measure, but in general we have attempted to analyze its influence through interviews with school staff that help us identify places where students feel strongly supported and places where students feel insecure. We have found that student academic achievement is highest when the bilingual/ESL staff at a given school feel very positive about the school environment, including the general level of administrative and teaching staff support and the context for intercultural knowledge-building provided for language minority students. This finding is reflected in other researchers' work (e.g. August & Pease-Alvarez, 1996; Lucas, Henze & Donato, 1990; McLeod, 1996; Moll, Vélez-Ibáñez, Greenberg & Rivera, 1990; Tharp & Gallimore, 1988).

In our study, certain school buildings were identified by bilingual/ESL staff as highly socioculturally supportive. Language minority students in these schools are respected and valued for the rich life experiences in other cultural contexts that they bring to the classroom. Their bicultural experience is considered a knowledge base for teachers to build on. The school is a safe, secure environment for learning. Native-English speakers treat language minority students with respect, and there is less discrimination, prejudice, and open hostility. Often sociocultural support includes an additive, enrichment bilingual context for schooling, where students' L1 is affirmed, respected, valued, and used for cognitive and academic development. Sometimes native-English speakers choose to join the bilingual classes, and both groups work together at all times in an integrated schooling context. In general, we have found that the school buildings with the strongest sociocultural support for language minority students are those that produce student graduates that are among the highest academic achievers in each school district.

Integration with the Mainstream

Cost-effectiveness and the duplication of existing services are issues that greatly concern every school administrator. Do all language minority students need add-on services, or can effective support be provided in grade-level classes? We have found that bilingual/ESL program models that find ways to integrate with grade-level classes in the mainstream instructional program can be highly effective, if they are carefully planned and implemented by well-trained bilingual/ESL school staff.

The curricular mainstream for native-English speakers serves several important functions for bilingual/ESL staff. For natural second language acquisition to occur, ELLs need access to meaningful interaction with native-English-speaking peers in a supportive environment. Same-age peers are a crucial source of L2 input. But English-speaking peers are only beneficial in a social setting that brings students together cooperatively (Wong Fillmore, 1991), including interactive negotiation of meaning and equally shared academic tasks. The teacher serves an important role in structuring the class tasks so that the L2 acquisition process is enhanced, and teachers need to be well trained to provide the sociocultural support for all students. Teaming of bilingual/ESL staff with grade-level teachers is one strategy used in some of the schools in our research sites. Administrators include extensive planning time in the school schedule when team teaching is in place. Ongoing staff development for all teachers is another important strategy.

A second function of the curricular mainstream is to continue the cognitive challenge. Student groups who are separated from grade-level classes for most of the school day for several years do not know the level of cognitive and academic work expected in the mainstream, and with time, students may develop lower aspirations for their own academic achievement. Ability grouping and tracking can lead to segregated patterns that do not provide students with equity and access to the full curriculum (Oakes, 1985, 1992). The schools in our study with higher academic achievement have eliminated most forms of ability grouping and tracking. They have found meaningful ways to school students together, for at least half of each school day, and in some special programs, for the whole day.

In our study, the program with the highest long-term academic success is two-way bilingual education. This is an integrated form of bilingual education in which all students may participate. Since this is a mainstream, grade-level model of schooling, it is the most cost-effective model of bilingual education, because add-on services do not need to be provided by extra staff. We will examine this model in the next section, where we illustrate the five program characteristics just discussed as they apply to some common program models in bilingual/ESL education.

Language Minority Students' Academic Achievement Patterns

To examine the long-term perspective from kindergarten through 12th grade, we have examined cohorts of similar students, following them for as long as they remained in each school system. The following figures presented in this publication represent general patterns of language minority student achievement across our five school district sites. Each line in each figure represents the typical academic performance of students across our five school district sites on standardized tests in English reading. This is the most difficult test of all, as it correlates strongly at the 11th grade level with the reading test of the SAT, a college entrance exam. The reading test measures problem-solving and thinking skills across the curriculum. In our findings, patterns of student performance on the standardized tests in science and social studies fall into the same general pattern as the English reading test. Mathematics and English language arts achievement of language minority students is slightly higher than their performance on the English reading, science, and social studies tests, but the same general pattern of performance, as well as the same ranking of long-term achievement influenced by program participation, is present in the mathematics and language arts data.

In general, when examining the two curricular areas on the standardized tests that focus on the English language--reading and language arts--we have found that the language arts tests tend to measure more easy-to-teach discrete-point skills; whereas the reading tests involve more complex problem-solving across the curriculum. The reading test is thus the most demanding--the "ultimate measure"--of all the curricular subtests of the standardized tests. When English language learners are able to reach age-appropriate grade-level norms on the reading subtest, and sustain that level of achievement in subsequent years, they have demonstrated that they can compete successfully with native-English speakers on the most difficult test given in school. More importantly, this is an indicator that they are moving toward true long-term educational parity with native-English speakers in all subjects, the ultimate goal for educating English learners.

The Influence of Elementary School Bilingual/ESL Programs on ELLs' Achievement

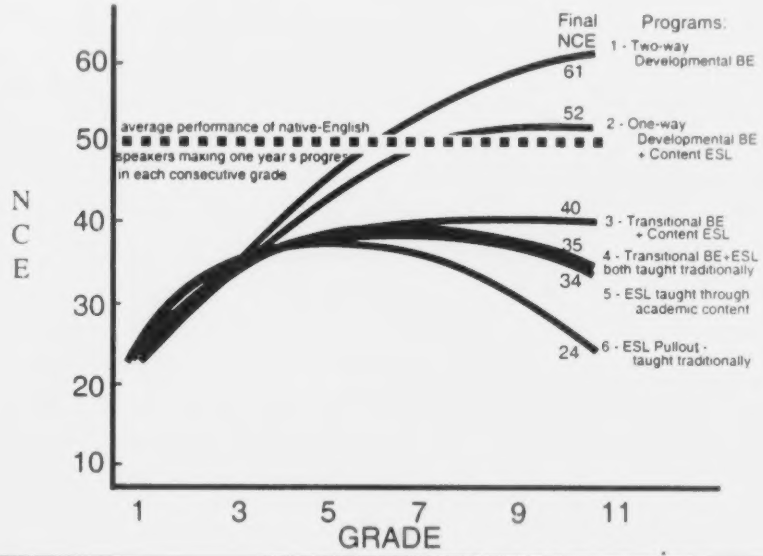
Figure 6 presents the patterns of the academic achievement of **students who begin schooling in the U.S. in kindergarten with no proficiency in English.** These students do not remain English language learners throughout their schooling, but they are all ESL beginners when they enter U.S. schools in kindergarten. It is important to remember that this figure represents cohorts of students who start school with the same general background characteristics--i.e., no

Figure 6

PATTERNS OF K-12 ENGLISH LEARNERS'
LONG-TERM ACHIEVEMENT IN NCEs
ON STANDARDIZED TESTS IN ENGLISH READING
COMPARED ACROSS SIX PROGRAM MODELS

(Results aggregated from a series of 4-8 year longitudinal studies from well-implemented, mature programs in five school districts)

Program 1: Two-way developmental bilingual education (BE)
Program 2: One-way developmental BE, including ESL taught through academic content
Program 3: Transitional BE, including ESL taught through academic content
Program 4: Transitional BE, including ESL, both taught traditionally
Program 5: ESL taught through academic content using current approaches
Program 6: ESL pullout--taught traditionally



BEST COPY AVAILABLE

proficiency in English and low socio-economic status as measured by eligibility for free or reduced lunch. Middle-income students with no proficiency in English were a separate group of student cohorts. We found this same general pattern of academic achievement in each major group of student cohorts (grouped by socioeconomic status) that we analyzed separately.

To create Figure 6, following English language learners' progress across time, we analyzed records of language minority students who attended each school district and were tested from 1982-1996. From over 700,000 student records, we were able to identify 42,317 student records in 4-year, 5-year, 6-year, and so on up to 8-year overlapping testing cohorts to present a longitudinal perspective. Each line thus has an underlying long-term longitudinal cohort, with a series of overlapping shorter-term longitudinal cohorts, confirming the general longitudinal pattern. Each line in the graph is defined by a weighted average of all of the cohort scores available at each grade level.

Each solid line in Figure 6 represents English language learners who received one type of program during the elementary school years only. Following these ESL beginners' participation in a special program at their elementary school (which could be a minimum of two years in Grades K-1--e.g. ESL pullout--to a maximum of seven years in Grades K-6--e.g. developmental bilingual education), all of these language minority students continued in the mainstream in grade-level classes with instruction all in English (L2) throughout their middle and high school years. Thus, number of years of instruction is considered as a part of the program's definition, and not as a variable to be controlled.

This is so because the instructional intent is quite different from one program type to another. For example, ESL pullout programs are designed to be short-term, limited instructional support from an ESL-certified teacher for a portion of each day. Thus, there are no 4-year, 5-year, or 6-year ESL-pullout programs in existence in our participating school systems. Since ESL-pullout programs address only one Prism Model dimension, the Linguistic area (and then only in English), and do not explicitly provide for students' continuing age-appropriate development in cognitive and academic areas while they are learning English, it is instructionally desirable that students have shorter exposure to such programs. Continued exposure to such an instructionally limited program would almost certainly produce larger gaps between English learners and native-English speakers with more years of this type of instruction, since students' cognitive and academic needs would be unaddressed for a longer period of time.

In contrast, developmental bilingual programs are designed to allow the students to continue age-appropriate development in all school subjects and to maintain native-speaker-like rates of cognitive development through L1 instruction while they are acquiring academic English. Thus, there are no 1-year, 2-year, 3-year, or 4-year developmental bilingual programs in our participating school systems. By definition, this program is long-term and addresses all of the Prism Model dimensions, rather than only one or two as in other program types.

ESL pullout (Line 6) was the most common program in our school district sites; 51 percent of the students in our sample attended this program. ESL content (Line 5) was attended by 13 percent of the students in the early grades; we found this a more common program in secondary schools.

Transitional bilingual education taught traditionally (Line 4) was represented by 17 percent of the student sample; while transitional bilingual education taught with more current approaches (Line 3) was represented by 9 percent of the sample. Seven percent of the students had attended a one-way developmental (or maintenance) bilingual program (Line 2), and three percent of the students had attended a two-way bilingual program (Line 1).

As can be seen in Figure 6, significant differences in student performance begin to appear as they leave their elementary school instruction and continue in the cognitively demanding secondary school years, with dramatic differences seen in student achievement by the end of their schooling. Yet when examined as a cumulative difference across ten school years, the difference between Line 1 and Line 6 is an average of 3.7 NCEs per year. That is, students attending the two-way developmental bilingual program were able to gain 3-4 NCEs per year more than typical native-English speakers. In contrast, students attending ESL pullout gained an average of zero NCEs per year over the 10 years, keeping pace with but not closing the initial achievement gap with native-English speakers.

These differences can be clearly attributed to program type attended in elementary school, since we took great care to match student cohorts by socioeconomic status, L1 and L2 proficiency, and amount of formal schooling, with all students in this longitudinal picture having received all their schooling in the U.S. Although other variables might exist on which these groups could be blocked or matched, our preliminary analyses indicated that these variables had the strongest effect on student achievement. After the effects of these variables were accounted for, further blocking, matching, or covariance adjustments with weaker variables resulted in non-significant adjustments, and were abandoned as ineffectual in subsequent analyses.

In Figure 6, the dotted flat line at the 50th percentile or NCE across Grades K-12 represents typical native-English speakers' performance on these tests, making 10 months of progress with each 10-month year of school. This is the national comparison group with whom English language learners are competing as they move through the school years. Our goal as educators is that students just beginning development of the English language, who therefore start school not on grade level in English (where grade level is defined as the 50th NCE on the tests in English), will as a group eventually reach the 50th NCE and be able to sustain that general level of achievement. On the local level, English language learners may also be compared to the local distribution of native-English speakers' scores.

To understand Figure 6, it is necessary to define some of the basic differences between programs. First, it is important to remember that **these data analyses present a historical picture of programs that existed during the 1980s and early 1990s.** These models have continued to evolve, along with the school reform movement of the 1990s, and still more variations exist today. Therefore this data provides information on some of the major program variations to date, but does not yet include all possibilities. This data represents findings from bilingual programs implemented only in the elementary school, including two-way and one-way bilingual 50-50 models and transitional bilingual education, following students after they left these programs. English-only approaches analyzed in this dataset include content ESL (also referred to as sheltered instruction and structured immersion) and ESL pullout. Future analyses that we will be conducting include data on two-way bilingual immersion 90-10 programs, as well as bilingual instruction at the secondary level.

To define basic program differences as represented in Figure 6, we have found that the five program characteristics defined in the previous section are important to examine--the amount of L1 support, type of L2 support, general types of instructional teaching styles, sociocultural support, and integration with the curricular mainstream. These differences were identified through our focus interviews with the bilingual/ESL staff in each school district.

Amount of L1 support. How much the primary language (L1) of the students is used for instruction is one of the most prominent characteristics that defines differences among programs in bilingual/ESL education. Figure 6 illustrates dramatic differences in long-term academic achievement, by the amount of L1 instructional support provided for language minority students in their elementary school program. **The more L1 academic work provided, the higher their achievement in the long term.** Lines 1 and 2 illustrate programs that provide a half-day in L1, taught across the curriculum, for Grades K-5 or K-6. Lines 3 and 4 illustrate programs that provide a half-day in L1 for Grades K-1, gradually increasing English instruction until L1 is phased out of the curriculum by Grade 4. Lines 5 and 6 illustrate programs that teach only through English, with no L1 support. To understand some of the detail of program implementation, two levels of decisions must be made regarding L1 instructional support: (1) how much of each school day or week instruction is provided in L1 (including which subjects or themes will be taught in L1 and which ones in L2) and (2) for how many years L1 instruction is continued (including the proportion of L1 instruction for each school day or week in each succeeding year).

Figure 7 provides short descriptions of common program models by defining the amount of instructional support for LM students' L1, beginning at the top of the figure with programs that provide the most L1 support and ending with those providing no L1 support. The reader must keep in mind that **all bilingual program models include English content teaching for at least some portion of the school day or school week.** In the figure, the primary language (L1) of language minority students is labeled the "minority language." English is the majority language. These terms, "majority language" and "minority language," clear up the confusion caused by the terms "L1" and "L2" when referring to two-way bilingual classes for two language groups acquiring each others' languages through the academic curriculum.

The first three program models listed in Figure 7--**bilingual immersion** (sometimes referred to as "dual language" or simply "immersion"), **two-way, and developmental bilingual education—are very similar in program characteristics, providing very strong support for L1 academic and cognitive development for language minority students for as many years as possible.** Our only reason for listing them separately is that they developed under different historical circumstances, and they can be different depending upon whether they are designed as two-way or one-way bilingual programs.

The distinction between one-way and two-way bilingual instruction was first made by Stern (1963). In **one-way bilingual education,** one language group is schooled bilingually. **Two-way**

bilingual education refers to an integrated model in which speakers of each of two languages (e.g., Spanish speakers and English speakers) are placed together in a bilingual classroom to receive instruction across the curriculum through both of their two languages. Two-way is a grade-level, mainstream bilingual program, since native-English speakers are included (no separation or tracking), and the class receives age-appropriate schooling across the curriculum.

The **immersion** model was originally developed in Canada in the 1960s for majority language students to receive their schooling through two languages—French and English—throughout their schooling, Grades K-12. As this model has been adopted in the U.S., it has become known as the 90-10 bilingual immersion model, implemented in most U.S. schools as a two-way program. It requires the strongest long-term commitment to academic development of the minority language along with the majority language. The 90-10 model requires initial emphasis on the **minority** language, because this language is less supported by the broader society and thus academic uses of this language are less easily acquired outside of school. By Grade 6, students have generally developed deep academic proficiency in both languages and they can work on math, science, social studies and language arts at or above grade level in either language. In research studies on this model, in both Canada and the U.S., academic achievement is very high for all groups of students participating in the program, when compared to comparable groups receiving schooling only through English (Cummins & Swain, 1986; Dolson & Lindholm, 1995; Genesee, 1987; Lindholm, 1990; Lindholm & Aclan, 1991; Lindholm & Molina, in press). **Our data presented in Figure 6 includes only two-way 50-50 programs**, because we did not yet have data from 90-10 programs for a longitudinal look across all the grades. **We are now receiving new data from 90-10 two-way schools, and this model looks even more promising than the 50-50 model.** But we will address this in future reports.

To avoid confusion, **it is important to understand the distinction between immersion education in Canada, and the program model labeled "structured immersion."** Immersion educators in Canada developed immersion to be the strongest form of bilingual education, providing a full commitment to schooling in two languages throughout Grades K-12. Initially during the first two grades (K-1) of an immersion program, students are "immersed" 90 percent of the day in the **minority** language, or the language less supported in the societal context outside of school. The promoters of structured immersion misinform educators when they state that it is based on the Canadian model. **In fact, structured immersion is the reverse of the Canadian model**, with no instructional support for the minority language and all instruction only in English, the majority language. **Structured immersion as it has been implemented in the U.S. is another form of content ESL, taught in a self-contained classroom, with instruction all in English.** In Figure 6, Line 5 demonstrates the performance of ELLs in content ESL and structured immersion programs that use current approaches to teaching; while Line 1 demonstrates ELLs' performance in 50-50 two-way immersion programs, also using current approaches to teaching. ELLs assigned to structured immersion programs, taught using highly structured materials that introduce students step-by-step to the English language, do even less well than the student performance in Line 5, achieving

Figure 7

PROGRAM MODELS IN BILINGUAL/ESL EDUCATION IN THE U.S.

(Ranging from the most to the least instructional support through the minority language)

Bilingual Immersion Education (also referred to as Dual Language Education): Academic instruction through both L1 and L2 for Grades K-12. Originally developed for language majority students in Canada; often used as the model for two-way bilingual education in the U.S.

- **The 90-10 Model** (in Canada, referred to as early total immersion):
  Grades K-1: All or 90% of academic instruction through **minority** language (literacy begins in minority language)
  Grade 2: One hour of academic instruction through **majority** language added (literacy instruction in majority language typically introduced in Grade 2 or 3)
  Grade 3: Two hours of academic instruction through majority language added
  Grades 4-5 or 6: Academic instruction half a day through each language
  Grades 6 or 7-12: 60% of academic instruction through majority language and 40% through minority language.

- **The 50-50 Model** (in Canada, referred to as partial immersion):
  Grades K-5 or 6: Academic instruction half a day through each language
  Grades 6 or 7-12: 60% of academic instruction through majority language and 40% through minority language.

Two-Way Bilingual Education: (This is not really a separate model, but a variation of bilingual immersion and developmental bilingual education.) Language majority and language minority students are schooled together in the same bilingual class, and they work together at all times, serving as peer teachers. Both the 90-10 and the 50-50 are two-way BE models. Developmental bilingual education, a funding category in the federal Title VII legislation from 1984 to 1993, can also be a two-way program.

Developmental Bilingual Education (historically referred to as Maintenance Bilingual Education; another term used by researchers is Late-Exit Bilingual Education): Academic instruction half a day through each language for Grades K-5 or 6. Ideally, this type of program was planned for Grades K-12, but has rarely been implemented beyond elementary school level in the U.S.

Transitional Bilingual Education (also referred to as Early-Exit Bilingual Education by researchers): Academic instruction half a day through each language, with gradual transition to all-majority language instruction in approximately 2-3 years.

English as a Second Language (ESL) or English to Speakers of Other Languages (ESOL) Instruction, with no instruction through the minority language:

- Elementary education:
  - **ESL or ESOL academic content, taught in a self-contained class** (also referred to as Sheltered Instruction or Structured Immersion--varies from half-day to whole-day)
  - **ESL or ESOL pullout** (varies from 30 minutes per day to half-day)
- Secondary education:
  - **ESL or ESOL taught through academic content** (also referred to as Sheltered Instruction --varies from half-day to whole-day)
  - **ESL or ESOL taught as a subject** (varies from 1-2 periods per day)

Submersion: No instructional support is provided by a trained specialist. This is NOT a program model, since it is not in compliance with U.S. federal standards as a result of the Supreme Court decision of *Lau v. Nichols*.

© Copyright Virginia P. Collier & Wayne P. Thomas, 1997

BEST COPY AVAILABLE

at or below the level of Line 6. (See Collier, 1992, for a synthesis of other published research studies on program effectiveness with NCE comparisons across programs.)

The term **developmental bilingual education** was first introduced in the U.S. in the 1984 Title VII federal legislation. This term emphasizes the students' ongoing linguistic, cognitive, and academic developmental processes in both L1 and L2 that are stimulated and strengthened in an enrichment bilingual model of schooling. **In Figure 6, we have adopted this term, developmental bilingual education, to represent all enrichment models,** which go by varying names—immersion, bilingual immersion, dual language, two-way bilingual, maintenance, and late-exit bilingual education. **Enrichment models of bilingual schooling generally contrast greatly with remedial models of schooling.** When the underlying goal of a program is to "fix" students who are perceived as having a problem, the program generally separates the students from the mainstream and works on "remediation." The consequence is usually that students receive less access to the standard curriculum, and the social status quo is maintained, with underachieving groups continuing to underachieve in the next generation. **When the focus of the program is on academic enrichment for all students, with intellectually challenging, interdisciplinary, discovery learning that respects and values students' linguistic and cultural life experiences as an important resource for the classroom, the program becomes one that is perceived positively by the community, and students are academically successful and deeply engaged in the learning process.**

Developmental bilingual programs use the LM students' L1 for academic enrichment for as many years as possible, teaching the school curriculum through L1 for a minimum of half a day for all of the elementary school grades, and continuing when possible into the middle and high school years. **Our findings illustrate that this strong L1 cognitive and academic development for the first 6-7 years of schooling provides the knowledge base needed for LM students who begin U.S. kindergarten with no proficiency in English to reach and maintain academic success in English throughout the secondary school years.** As can be seen in Figure 6, without this level of L1 support, language minority students lose ground with each passing year as they reach the cognitively difficult years of high school, when compared to typical native-English speakers' continued academic growth during this period.

Type of L2 support. Programs can differ dramatically in the way the English language is taught. The major difference that influences student achievement is whether academic content in all school subjects is taught by ESL-certified teachers or whether the focus in ESL lessons is solely on learning the English language. We have found that programs that teach ESL through academic content (taught by ESL-certified teachers who understand second language acquisition) help students to gain an additional 10 NCEs (by the end of schooling) beyond the achievement level of peers who receive ESL focused only on the structure of the English language (ESL pullout, taught traditionally), as illustrated in Lines 5 and 6 of Figure 6. It should be noted that a 10-NCE difference between these two groups is equivalent to about one-half of a national standard deviation, a very significant difference both statistically and practically, in favor of ESL taught through content.. In ESL pullout programs, students do receive academic content taught by the mainstream teacher in English, but our data show that students do not benefit from the mainstream instruction nearly as much as they do when they receive academic content taught by an ESL-certified teacher who also has formal licensure to teach the content areas.

This provides a good example of clear differences among programs in their prevailing practices for L2 support that can influence student achievement. In both ESL pullout and ESL content programs at our research sites, students received special instructional support from an ESL teacher for the same general number of hours per day, gradually moving to all mainstream classes towards the end of their third year of development of the English language. Therefore, the same amount of instructional time spent with students of similar needs, but different teaching goals, leads to dramatic differences in student achievement across the years. Teaching ESL through academic content, with simultaneous language and content objectives, is clearly superior to limiting the focus of ESL to teaching the structure of the English language. We conclude this because content ESL classes resulted in an average achievement gain of 10 NCEs in the long term, when compared to ESL pullout classes.

Other examples of program differences by the type of L2 support associated with each program can be seen in our research findings illustrated in Figure 6. Students who attended either two-way or one-way developmental bilingual classes during their elementary school years (Lines 1 and 2) received strong academic content in English for a half-day, taught by ESL-certified staff. Both groups demonstrated long-term academic success in English, when strong academic content was provided in both L1 and in English, with each language given equal importance in the curriculum. Line 3 illustrates student performance in transitional bilingual education (TBE), in which ESL-certified staff taught ESL through academic content during the English portion of the school day. Students reached higher long-term academic achievement levels in these content-ESL classes than in TBE taught traditionally, in which the ESL teacher placed greater emphasis on teaching the structure of the English language (Line 4).

Type of teaching style. This program variable is closely connected to the one just discussed. Classes that are very interactive, in which the teacher facilitates discovery learning across all the curricular subjects, enhance the learning process, resulting in higher LM student achievement. LM students make less academic progress in passive, traditionally taught classes. In our research sites, ESL content was typically taught through a more interactive, interdisciplinary, discovery learning approach; whereas ESL pullout teachers tended to describe their focus as limited to teaching English structures, pronunciation, and vocabulary, oral and written, with any support for students' academic work in other subject areas as tutorial in nature, one-on-one with the ESL pullout teacher. This pattern resulted in a fairly traditional style of teaching, and not one that the ESL pullout teachers necessarily preferred. They described frustrations with their limited time with students and the difficulty of using cooperative learning and other more innovative approaches to teaching when ESL students of varying ages came and went from their classes at all times of the day, making it difficult to go in depth into content lessons.

In all of the program models, when a majority of teachers described their classes as very interactive, with a focus on interdisciplinary problem-solving, and making use of the students' knowledge and resources from their diverse life experiences in other linguistic and cultural contexts, students reached a higher long-term level of academic achievement, resulting in a cumulative 8-10 NCE gain by llth grade. In our interviews, teachers were the clearest about their "current approaches" to teaching in two-way and one-way developmental bilingual classes. These teachers perceived their role as facilitating a challenging, grade-level class across the curriculum, and they enjoyed watching their students "take off" when they became deeply engaged in the learning process, helping each other learn, and tackling difficult academic work with confidence, in either L1 or L2.

Sociocultural support. The six lines of Figure 6 demonstrate a hierarchy of program differences for two important variables--L1 academic instruction and sociocultural support, from the most support illustrated in Line 1, to the least support illustrated in Line 6. Two-way bilingual programs (Line 1) provide the most sociocultural support for LM students because the native-English-speaking students in the school choose to participate in the bilingual classes, thus indicating a willingness to work with and become friends with the LM students. With time, their joint participation in a cooperative learning setting tends to bring the two groups to respect and value each others' knowledge and to serve as peer tutors for each other. They stimulate each other cognitively as they learn together, and areas of potential cultural or linguistic conflict are openly dealt with and resolved as they become collegial collaborators in a curriculum with a more global perspective, reflecting the linguistic and cultural diversity of the participants. Instead of hostility, discrimination, suspicion, and prejudice expressed among students who do not work together, the students in two-way bilingual classes grow to value and respect each other in this shared learning environment. Teaching staff likewise affirm both groups' primary languages and celebrate bicultural school experiences that enrich and expand the cognitive challenges that come from intercultural knowledge and growth. The emotional side of learning is addressed as teachers understand and affirm students' differences as strengths and resources for the classroom. Teachers create a sociocultural support system in the classroom that gives students the emotional security they need to accelerate the learning process. The overall result is that the LM students enjoy the same favorable sociocultural environment for learning all school subjects (including English) as is normally enjoyed by native-English speakers.

This sociocultural support is also present in a one-way developmental bilingual program, even when native-English speakers do not participate in the bilingual classes (Line 2). In this context, the bilingual and ESL teachers provide strong emotional support, understanding and affirming LM students' bicultural needs, and providing the curriculum through both L1 and L2 for enough years that the academic gap in L2 is closed, giving students the academic takeoff point where they can then stay on grade level in L2 even when the curriculum gets more cognitively difficult at secondary level. The key is for students to develop a strong self-identity and comfort level regarding their bilingual/bicultural heritage in the early grades, which helps them to stay on grade level in their

L1 throughout the elementary school years while they are closing the gap in L2. Then, as adolescents moving into middle school, they have already developed the sociocultural support system within themselves, as well as the cognitive, academic, and linguistic strengths that they need to stay on grade level in L2 throughout the remainder of their schooling. In doing so, they also tend to keep pace with the constantly advancing native-English speakers instead of falling behind by a small amount each year that cumulatively becomes a large achievement gap by the end of their school years.

Transitional bilingual programs (Lines 3 and 4) also provide very strong sociocultural support. The only problem is that the support is not provided long enough for LM students' long-term success. The sociocultural support and L1 academic support is taken away too soon when LM students are moved into the mainstream in fourth grade. When it is time to exit the transitional program, since the LM students are not yet on grade level in L2 (as is true of all programs by fourth grade--see Figure 6), even though they make good progress with each succeeding school year (as much as the average native-English speaker makes), they are not able to do what is necessary to completely close the gap. To do this, the typical LM students must outperform the typical native-English speakers for several consecutive years to close the initial achievement gap and reach the same average score as the native-English speakers, the 50th NCE. Similarly, ESL pullout and ESL content teachers who are cross-culturally sensitive and have been well trained may provide strong sociocultural support for the LM students while they attend the ESL classes, but the remainder of the school day is in the mainstream where they generally have much less sociocultural support from peers and teachers.

Integration with the curricular mainstream. This program variable can be thought of as a fine balancing act. Separate classes for part of the school day can serve a very important function for many language minority students. English language learners need teachers who understand the long-term process of second language acquisition and can provide them access to both language and academic content through L2 in the first years of schooling in L2 , so that the curriculum more carefully meets their unique needs. Given a talented, caring, experienced mainstream teacher who can handle heterogeneous classes with students who vary greatly in their proficiency in English, it is possible that ELLs can benefit from a half day in a mainstream class (with the other half of each school day in instruction in L1). But not all mainstream teachers have the staff development training and experience to provide for appropriate curricular needs of the English language learner. ESL teachers who can teach language through academic content serve the crucial function of providing appropriate and meaningful access to the academic curriculum for students in their first three years of development of the English language.

At the same time, ESL students who are separated from the curricular mainstream for many years have no knowledge of academic expectations within the mainstream. When they face the reality of the academic achievement expected of native-English speakers of their same age, if they have been kept in a program that watered down academic content and did not allow them to make the leaps needed to close the academic gap, they may never catch up to the constantly-advancing native-English speakers. Furthermore, kept in isolation from native-English-speaking peers, they do not receive access to crucial peer input in L2 for the natural second language acquisition process to be stimulated to its highest level.

However, the "fine balancing act" between the mainstream and separate classes becomes still more complicated when L1 instruction is considered. A crucial function of separate classes for a portion of the school day is to provide LM students on-grade-level academic instruction in their primary language, so that they can keep up with cognitive and academic development appropriate to their age group while they are acquiring deep academic proficiency in English. Academic work taught through L1 and L2 can be provided through two-way, integrated bilingual classes in the curricular mainstream, but this is possible only where the L1 of a given LM group is a language that the native-English-speaking parents wish for their children to learn. Furthermore, some schools choose to continue one-way bilingual classes for the LM students of one language background when there is a shortage of bilingual teachers who are academically proficient in the minority language of that LM group. Administrators in these schools explain that given that bilingual schooling cannot yet be provided for both language groups, it is more important to assist the LM students first when their achievement is lower than that of the native-English-speaking students. When the bilingual teacher shortage is resolved, two-way bilingual schooling can be provided for all students who choose to be in this type of enrichment, mainstream program.

In Figure 6, Lines 1 and 2 provide an interesting contrast, as an illustration of the complexity of program decisions regarding integration into the curricular mainstream. **Line 1 illustrates a bilingual program that is a mainstream program,** in which students are not separated into "remedial" classes for any period of time. LM students who are just beginning development of the English language are a part of this program from the moment they enter the school system, including new immigrants who have just arrived in the U.S. and are placed in their age-appropriate grade level. Bilingual and ESL teachers in two-way bilingual programs learn to adapt to teaching very heterogeneous classes. Teachers depend on the two student groups to serve as models for L2 input, through peer teaching in a discovery learning classroom, which activates cognitive and linguistic development in each language. New immigrants who have received strong schooling in their home country can serve as excellent peer models of the minority language of instruction, and native-English speakers in the same class serve the same function in English.

Line 2 illustrates LM students' long-term academic achievement in a one-way bilingual program. Depending upon how this type of education is viewed, it could be perceived as a program separate from the mainstream. For example, student achievement could be less than optimal if the bilingual/ESL staff do not collaborate with other teaching staff to make sure that the work in L1 is on-grade-level, and that the work in L2 is gradually catching students up to grade level, as they move through the elementary school years. However, in our data examining LM students' achievement in five school districts with very experienced staff, we found that these one-way bilingual programs that take seriously the responsibility for continuing academic development through L1 throughout the elementary school years, are able with time to close the achievement gap in English, so that by the time LM students leave their L1 instruction and move into middle school, they are on grade level in English and succeed in maintaining grade-level achievement throughout the remainder of their schooling. As bilingual staff in these programs described their teaching practices, they saw themselves as collaborators with all staff in the school system, aware of grade-level expectations, and prepared to assist their students to achieve to the same high levels as expected of all students. Furthermore, as ELLs in these programs reached higher levels of proficiency in English, bilingual/ ESL staff worked collaboratively with all teaching staff to integrate LM students appropriately with native-English-speaking students during the English portion of the school day. Thus, **while a one-way bilingual program may not integrate LM students into the native-English-speaking mainstream classroom, bilingual/ESL staff can successfully guide LM students toward success in the curricular mainstream**. In fact, LM students attending programs represented in Lines 1 and 2 are the high achievers in the long term, slightly outperforming (Line 2) and even strongly outperforming (Line 1) typical native-English speakers on the most difficult tests given in school (with the typical performance of native-English-speaking students represented by the 50th NCE).

The remaining programs present this same delicate balance between separate schooling and integration with the mainstream. Some separate schooling appears beneficial--especially L1 academic instruction as well as ESL content--when mainstream classes do not meet LM students' needs. But a watered-down curriculum, as well as less access to cognitive and academic development in L1, does not provide students with the cognitive push that they need to close the achievement gap with the native-English speaker. Programs represented by Lines 3 and 4 are mostly classes separated from the mainstream until Grade 4. Lines 5 and 6 represent programs that separate students from the mainstream for a portion of each school day for the first 2-3 years of schooling in the U.S., but the mainstream provides little accommodation for ELLs, with teachers mainly teaching to the needs of the native-English-speaking students.

Interaction of the five program variables. To summarize, **our research findings clearly illustrate the importance of providing strong on-grade-level cognitive and academic development through students' L1 for as long as possible.** ESL content programs provide the most effective L2 support, the appropriate teaching style, the sociocultural support while students are attending the ESL content classes, and integration with the curricular mainstream. But without L1 academic support, even when all of the other four program variables are provided, this is not sufficient assistance for English language learners eventually to close the academic achievement gap. In our research findings, ELLs who start school with no proficiency in English and receive a quality ESL content program have as a group reached the 34th NCE by Grade 11, but they are no longer closing the gap (see Figure 6). In contrast, students who receive strong L1 support throughout the elementary school years (Lines 1 and 2) are at the 50th NCE or above by Grade 11. Students who receive some L1 support until Grade 4 and the other four program variables are provided (Line 3) are able to reach the 40th NCE by Grade 11. **Of all the five program variables, L1 support explains the most variance in student achievement and is the most powerful influence on LM students' long-term academic success.**

The Influence of Secondary School ESL Programs on ELLs' Achievement

To analyze long-term academic achievement patterns of language minority students attending middle school and high school, we first interviewed bilingual/ESL school staff in our five school district sites to examine the needs of LM students and the types of programs provided for them at secondary level. We have already presented in the previous section the results of our analyses of secondary LM students who had attended U.S. schools since kindergarten. We found that special programs provided for LM students at secondary level that were in existence for at least ten years in our school district sites (to get the long-term perspective) were limited to basically two types of ESL classes--ESL taught through academic content and ESL taught as a subject. No bilingual instruction was provided at secondary level for enough years for us to examine its long-term effects. We are now receiving new data on secondary bilingual instruction and will provide analyses of that data in future research reports.

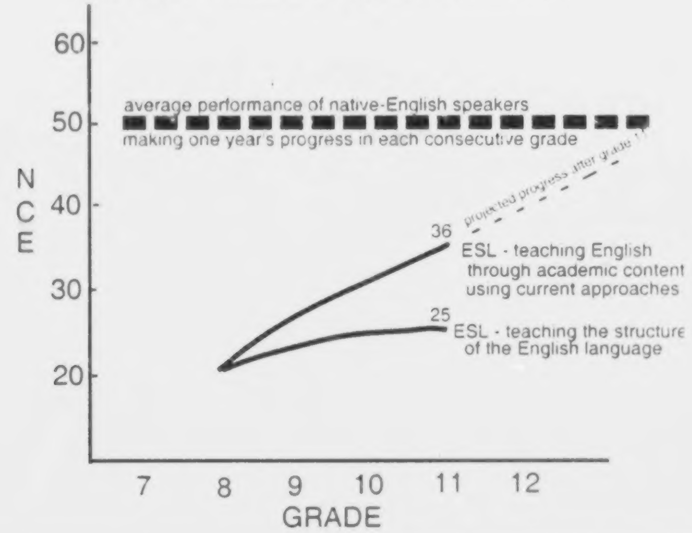
As we found with the elementary school achievement data, in our initial analyses of the secondary achievement data, LM students who were not proficient in English when they entered the school system took many years to close the academic achievement gap when compared to native-English-speaking students on grade-level tests in English. Since the standardized tests in English were typically administered at eighth and eleventh grade levels, and since it took several years of exposure to English before ELLs were allowed to take this type of test, we have chosen from all our data analyses of different cohort groups to present in this report the general pattern of achievement seen among students who arrived for the first time in U.S. schools in Grade 5 or Grade 6 and who had prior L1 schooling in their home country and were tested as on grade level in L1 when they arrived. Among the various groups defined by amounts of prior L1 schooling and degree of on-grade-level performance in L1, this group attained the highest level of achievement by the 11th grade. We found that L1 grade-level schooling in home country was an important predictor of academic success in L2, with those students who had experienced interrupted schooling achieving at a much lower level in L2 than that reported in Figure 8.

As can be seen in Figure 8, we are again reporting in NCEs, examining LM student progress as measured by the standardized tests in English reading, which correlate strongly with the SAT, used as an admissions criterion for 4-year undergraduate study at university level. To repeat this important point, the English reading subtest is the most difficult test given in school, as it measures problem solving across the curriculum. To do well on this test, students must make use of their math, science, social studies, language arts, and literature knowledge, applying this combined knowledge to curricular problem-solving. While we have examined LM students' performance on other measures used by the school districts in our study, such as criterion-referenced tests and performance assessments, we find the same general pattern in LM students' academic achievement across the varying measures. We are therefore using one type of measure--the English reading subtest of the standardized tests--to illustrate the typical pattern that we see in LM students' performance on several different types of measures. Remember that these are very difficult academic tests taken towards the end of high school, measuring complex cognitive

Figure 8

GENERAL PATTERN OF SECONDARY LANGUAGE MINORITY STUDENT ACHIEVEMENT ON STANDARDIZED TESTS IN ENGLISH READING FOR NEW IMMIGRANTS WITH PRIOR L1 SCHOOLING WHO ARRIVE IN THE U.S. IN GRADES 5-6

(Results aggregated from a series of 3-year longitudinal studies
from well-implemented, mature programs in five school districts)
© Wayne P. Thomas & Virginia P. Collier, 1997



and academic development at levels appropriate for native speakers of English.

As stated with Figure 6, the dotted line represents the comparison group of typical native-English-speaking students, who define the 50th NCE. In other words, this flat, dotted line represents the typical performance of students across the U.S. on these tests. The tests are cognitively more difficult with each year of school, because the tests at each subsequent grade represent the additional knowledge gained in one ten-month school year. The average students who score initially at the 50th NCE and then stay at the 50th NCE the following year--a zero NCE gain--are students who have made a full year's progress in a year's time. Students whose NCE scores go down from one year to the next have made less than a full year's progress in a year's time, thus widening the achievement gap with typical native-English speakers even further. Students whose NCE scores increase from one year to the next have made more than one year's progress in one year's time and are beginning to close the achievement gap with typical native-English speakers.

To eventually reach the typical academic performance for each grade level that U.S. students make, English language learners who begin their U.S. schooling with no proficiency in English must gradually move from the 1st NCE (assuming no guessing on the test) to the 50th NCE over time.

BEST COPY AVAILABLE

This means that they must make more progress with each year of school than the typical native-English speaker makes to ever close the academic achievement gap on school tests. When a line rises in Figure 8, this means that the students are making NCE gains with each year of school; in other words, they are making more progress than native-English-speaking students of the same grade level make. When a line stays flat, that means that the LM students are, at that point, making a full ten-months' progress in ten months' time, but they are not closing the gap with the constantly advancing native-English speaker.

As can be seen in Figure 8, we have not included ELLs' performance on this standardized test when they first arrived in the U.S. in the 5th or 6th grade, because this type of test is not given to beginning ESL students in English. Upon arrival, they can demonstrate what they know in L1 but not in L2. Our school district sites tested arriving students on an English language proficiency measure and on L1 reading and math measures when possible. In their first three years of U.S. schooling, these students received one of two types of ESL programs--ESL taught through academic content, or ESL taught as a subject, in which the focus was on learning the structure of the English language. The ESL support was provided for 2-3 academic periods per day, with the remainder of their day in mainstream classes. By the end of eighth grade, these fifth and sixth grade arrivals, well schooled in their home countries for Grades K-4, had been exited from ESL and for the first time were given the standardized test in English. After 3-4 years of exposure to the English language in the U.S. along with well-implemented ESL classes taught by experienced ESL staff, the typical performance of groups of former ELLs had moved from beginning level (1st NCE) to the 20th NCE on this cognitively difficult standardized test. This is remarkable progress in 3-4 years. It is extremely unrealistic for policy makers to expect these students, as a group, to be at the 50th NCE (i.e. on a par with typical native-English speakers of the same age) in only one or two years, on this type of difficult academic measure. All of our research findings consistently demonstrate the extensive length of time involved in the developmental processes of L2 schooling, because these tests are measuring not only English language development but also cognitive and academic growth.

As these students continued in high school, the standardized test was given for the last time at the 11th grade level. What we found at this point was a very significant difference between the achievement of those LM students who had attended ESL content classes compared to those who had attended ESL classes that provided only English language development. Both programs were comparable in the amount of time devoted to special support and the experience and competence of the ESL teachers. But students who had been taught both ESL and academic content simultaneously were making the gains needed with each succeeding school year to eventually close the gap with native-English speakers' achievement, closing the gap at the rate of about 3-4 NCEs per year. LM students who had received well-taught ESL classes focused on the structure of the English language and the remainder of their coursework taught by mainstream teachers were making only 1-2 NCE gains with each year and remained among the lower achieving students in the U.S. (in the bottom one-eighth) by the end of their schooling. In contrast, the LM students who had received an ESL content program demonstrated consistent NCE gains with each year of school, sufficient to show that their projected progress after grade 11 would get them to the level of typical native-English speaker performance by freshman or sophomore year of college. By 12th grade, these LM students were sufficiently high achieving to gain admission to a four-year university.

In interviews with ESL teaching staff at secondary level in our research sites, teachers described the following characteristics as **essential to content-ESL program success: (1) teaching second language through academic content, (2) consciously teaching learning strategies needed to develop thinking skills, solve problems, and apply new knowledge, (3) activating and connecting students' prior knowledge (considered a class resource) to the new knowledge developed in class, (4) respecting and valuing students' home language and culture and using students' L1 at appropriate times for academic work in small groups, (5) using cooperative learning, (6) facilitating an interactive, discovery learning classroom context, (7) encouraging intense and meaningful cognitive and academic development (to make up lost time in academics while acquiring English), (8) assisting students with access to and use of technology, and (9) using multiple measures across time for ongoing classroom assessment.** These characteristics summarize what we would classify as "current approaches" to teaching ESL at secondary level.

School Leavers

We plan to present our data on dropouts (more recently referred to as "leavers") in future reports, but the following provides an overview of our findings to date. We have found that many language minority students do not complete high school. In our data, LM students who received ESL pullout with no L1 schooling are most likely to leave school before high school completion. Had those students stayed in school and been tested with their peers, we suspect that Line 6 of Figure 6 (LM students' achievement following ESL pullout) would show even lower academic achievement by the 11th grade, since these leavers' scores would almost certainly have been below the average for their group, thus lowering the existing group average score even further.

In our data, LM students who had attended a two-way or one-way developmental bilingual program in elementary school were the least likely to leave school. For new arrivals at secondary level, LM students who arrived with on-grade-level L1 schooling from their home country for at least Grades K-4, and who received a content ESL program as described above, were the least likely to leave school. LM arrivals who had experienced interrupted home country schooling were the most likely to leave before completing high school. In future data analyses we will examine bilingual schooling for secondary students and provide more detailed reports on school completion patterns.

PHASE II OF THIS STUDY: 1996-2001

Appendix B describes the next phase of this study currently in progress. Funded by the Office of Educational Research and Improvement of the U.S. Department of Education from 1996 to 2001, we have expanded the number of school district sites working collaboratively with us to collect and analyze data for the purpose of answering urgent questions posed by education policy makers in many regions throughout the United States. Our study is one of 30 studies being conducted by the Center for Research on Education, Diversity, and Excellence (CREDE), located at the University of California, Santa Cruz, and directed by Dr. Roland Tharp. We encourage you to watch for CREDE publications summarizing the results of these studies over the next five years.

RECOMMENDATIONS

The major intents of this publication have been:

- to describe the findings of our past ten years of data collection and analysis in many school districts around the country, with special attention to the five school districts with which we have worked for the past five years;
- to describe our findings in terms of the underlying theory that supports and explains them, the Prism Model of Language Acquisition for School;
- to test the Prism Model with our data analyses in order to make predictions about long-term student achievement that can be validated by other school systems and researchers;
- to make general policy recommendations to our readers in school systems around the country who wish to know how our findings apply to their local context;
- to summarize the action recommendations that we have made individually to each of our participating school systems, who provided us with access to the data analyzed in this long-term research program, and to summarize the recommendations that apply beyond these five school districts for use by other school systems.

Having completed the first two tasks listed above, we turn now to the third and fourth. In doing so, we wish to speak directly to school systems that are interested in constructively reforming their instructional programs for language minority students and that are ready and willing to take action immediately.

Policy Recommendations

Recommendation 1: Change your thinking regarding the goal of research and evaluation in language minority education. Be prepared to undertake long-term actions and to look for long-term results, while de-emphasizing short-term studies or program evaluations for school decision-making. Be prepared to ask better questions about program effectiveness.

For years, the prevailing research question has been defined by the thinking of short-term evaluation: "Which program for English language learners is better (leads to higher achievement) in the short term (1-2 years), controlling for initial differences between the students in each program?" After 25 years of politically charged assertion and counter-assertion, many researchers (including us) can agree that there are no substantial short-term differences among programs for English language learners, especially in the early years of schooling (Grades K-3).

In this report, we emphasize that it is necessary to look past the short-term view of English learners' experiences in the schools, to look past a focus on the early years of schooling to the final outcomes of schooling over many years, and to look beyond acquisition of English to mastery of the full curriculum. Instead of asking the short-term question, "Which program is better during the first 1-2 years?" we emphasize the long-term view of the schooling of English learners, as well as all language minority students, by asking the following refined and improved research questions:

> "Which instructional practices (considered as features of instructional programs for English learners) allow English learners to reach full parity in the long term with native-English speakers in mastery of the full academic curriculum? How long does it take for English learners to reach full parity so that, as a group, they are indistinguishable from native-English speakers by the end of their school years?"

An additional improved research question is somewhat similar to the traditional "Which program is best?" question, except that it avoids applying experimental research procedures that are inappropriate in a school-based field setting, provides for alternative means of controlling extraneous variables, follows students for the long term rather than only for the short term, and greatly increases the sample size (and thus the statistical power) of the study. It also avoids looking at 'typical' schools, many of which do not implement their programs well, whatever the program type used. Specifically, it focuses on a purposive sample of schools whose programs for English learners are well-implemented, whose teachers are experienced, whose programs are long-running and stable, and whose students are selected for their similarity in prior exposure to English, levels of formal schooling, and family socioeconomic status. It then follows these similar groups of students longitudinally for as many years as possible, documenting the degree to which they do, or do not, close the initial achievement gap with native speakers of English. Programs are considered better if their students close the gap over time to some extent, and are considered best if they allow typical English learners to completely close the achievement gap (and keep it closed thereafter) by the end of the school years. This research question might be stated as follows:

> "For students who begin instruction in kindergarten and continue their instruction for several years thereafter, and who are (1) tested on school tests in English only after 3-4 years in school when they can take these tests in English with some facility, (2) similar in prior exposure to English, (3) similar in family socioeconomic status, and (4) similar in number of years of formal schooling, what is the long-term 'high water mark' of student achievement that each major type of instructional program can be expected to produce by the end of the students' school years, when each program is well-implemented by fully trained teachers in good school systems?"

These are appropriate research questions for school-based staff to address, as a means of informing the program and policy decisions that they must make. The short-term questions provide little or no useful information for school-based decision-makers who are seeking the best programs and instructional practices for their language minority students.

Recommendation 2: Collect data that is both cross-sectional and longitudinal, and examine successive cohorts of students, in order to get the full picture of the effects of your instructional programs for English language learners, as well as for all language minority students.

The typical school system consists of students who have attended for 1 year, 2 years, 3 years, etc., up to 12 years. Thus, on a given day, there is a 1-year attendance cohort, a 2-year cohort, a 3-year cohort, etc., in each school. Looking back at records of past student performance, there are similar multi-year cohorts of students to be studied. Some school-based questions that focus on the present status of selected variables (e.g. attendance, disciplinary actions, current achievement levels) require cross-sectional data and can be addressed using a short-term outlook. However, the impact of appropriate education for English language learners requires a long-term look at trend data, and a continuous monitoring of the progress that students make over a number of years. For these questions, only longitudinal data will do. If you must take a short-term view, focus it on the outcomes that your students can demonstrate at the end of their school years, not the beginning.

Recommendation 3: Realize that you must embark on a long-term effort to improve the outcomes of your school's instruction in all subjects and for all students. Improving language minority students' performance is a long-term undertaking, even under the best and most favorable of instructional environments and programs.

Your language minority students who are not yet proficient in English do need to acquire English, but don't let them fall behind the constantly advancing native-English speakers both cognitively and in their academic subjects while they are learning English. If possible, allow all of your students to acquire both English and another language as part of their formal schooling. Remember that all humans acquire language (first and second) as part of a long-term developmental process that can be slowed down by inappropriate instruction, but that cannot be speeded up beyond the limits imposed by the physical, psychological, and emotional development of your students.

Recommendation 4: Determine the expected long-term achievement that will result from continued implementation of your present program for English language learners. Then, determine which program's long-term achievement corresponds best with your expectations for your students.

To do this, first look at Figure 6, which presents the long-term achievement of students in each of six major program types, each with its associated instructional features. Determine the program line that best fits your school's instructional practices for LM students. If your chosen instructional approach is well implemented over at least a five-year period, you can expect the long-term achievement for your students to be at or near the points on the figure for that program and for the appropriate grade of your students five years from now.

Now, choose the *best* instructional program that can be implemented in the "real world" conditions in which your school operates (you are the best judge of this) and note its long-term student achievement potential. If you can professionally accept that level of potential long-term average achievement for your LM students, then fully and completely implement that program and stay with it for the next 3-5 years. If you are unhappy with the long-term achievement potential of your present instructional program, then begin to consider the possibility of "moving up" to the instructional program with the next highest long-term achievement potential. As you move up the program lines, you will be choosing to implement more of the predictor variables from our research that are associated with higher long-term LM student achievement.

Recommendation 5: "Move up" to a well implemented instructional program for English learners whose long-term predicted achievement matches your expectations for your students.

If you are currently using a program with low long-term achievement potential (for example, ESL pullout traditionally taught), resolve that your school will "move up" at least one line to a better program on Figure 6 during the next three years. This means that you resolve that you will use well-designed and validated instructional strategies, train the teachers in their use, and monitor the implementation of the program for students who are now in elementary school. For secondary students, this means that you will compare your English language learners' achievement to that of native-English speakers, note the size of the achievement gap, and respond with instructional strategies that feature as many of our predictor variables as possible.

Recommendation 6: Resolve that you will faithfully and fully implement your instructional program of choice for 3-5 years and that you will follow student achievement in all content areas during this time. When your students are in their middle school years (e.g. Grade 8), your district will test them using a standardized norm-referenced, criterion-referenced, or performance assessment instrument, and you will compare the performance of former ELLs and other LM students to that of native-English speakers at this time, prior to the rigorous cognitive demands and advanced coursework of high school. If there is a significant achievement discrepancy between the three groups at this time, resolve to implement a secondary plan for all LM students that will help them close the achievement gap with native-English speakers before the end of high school. What form should it take? Consult our list of predictors and implement as many of them as possible in the instructional program that has your professional confidence that it will most improve the long-term achievement of your LM students.

Recommendation 7: Implement your chosen instructional practices as well as possible and monitor your instructional programs continuously, making sure that teachers know how to use the instructional strategies that you've agree to implement, and that appropriate resources are available for instruction. Monitor your students' achievement on an annual basis if possible, but at least on an every-three-years basis.

Recommendation 8: Ask yourself, "Have our present instructional practices created long-term parity for language minority students with native-English speakers?" Arrange for your school or your school system to take the Thomas-Collier test of equal educational opportunity, as described in the next section of this report. Verify for yourself that our results, especially those in Figure 6 and Figure 9 (in the next section), do indeed describe the long-term results that your students have experienced as well. We already have heard from three large school systems who have validated our findings in this way. After you have validated our findings to your satisfaction in your school system, re-affirm your choices as described in recommendations 4 and 5 above.

Recommendation 9: Close that achievement gap and keep it closed!

Your students deserve no less. Fret less about what is politically expedient. Stop worrying about how to compare programs with experimental precision, and be more concerned about what instructional practices (not programs), in your best professional judgment, will reduce the large achievement gap that presently exists between your language minority students and your native-English speakers. It is probable that many (if not most) of your language minority students were born in this country, and their rights as citizens include the right to equal educational opportunity in the form of full educational parity with their native-English-speaking peers. For this to occur, you the educator must investigate what's working and what's not working for English learners as they move through the school years. You must inquire as to the long-term outcomes of your instruction and be prepared to change your strategies and practices to achieve better long-term results for your students. You must be prepared to implement well your chosen instructional strategies, so that you can compare well-implemented alternatives, rather than poorly implemented ones.

Do the right thing, as your best professional judgment defines it, to assure that language minority students' success in school will lead to their becoming fully productive citizens. We'll need all the productive citizens that we can get in the 21st century! When today's baby boomers begin to retire in droves 15-20 years from now, your students will assume society's burdens. In our own personal enlightened self-interest, and in the interest of our nation in the early 21st century, let's make sure that by the year 2030, 40 percent of the nation's school-age population--our language minority students--will be ready.

How Is Your School System Doing? -- The Thomas-Collier Test

Is your school system allowing its English language learners to achieve parity in long-term achievement with native-English speakers? You can use the Thomas-Collier test of equal educational opportunity to find out. Here is how it works:

Step 1: Examine your district-wide test results (norm-referenced, criterion-referenced, or performance assessment) in the last grade in which you test your students. For example, let's assume that you administer a nationally-normed test in Grade 11, toward the end of your students' school years.

Step 2: Separate out the scores of all students who have attended your school system for five years or more; set aside the scores of those who have attended your schools for less than five years. Also do not include the scores of former ELLs who arrived in your school system in the upper grades with interrupted or no previous formal schooling.

Step 3: Separate the "five-year" groups into three subgroups: those who were previously English language learners (ELLs), those who are language-minority (LM) but not ELLs, and those who are native-English speakers and not ELLs or LM.

Step 4: Compute the average 11th grade test scores for each of the three groups. Use raw scores, NCEs, or scaled standard scores but do not use grade-equivalent scores or percentiles, since these are not equal-interval data and are misleading.

The Thomas-Collier test consists of the following comparisons using the above data:

If your instructional practices are effective for native-English speakers, LM students, and ELLs (e.g. if typical ELLs outgain the national norm group by about 5 NCEs per year), then your ELLs should have closed the initial achievement gap with native-English speakers in about five or six years. Is this the case? In other words, if your instructional practices have been effective, then former ELLs should have closed the achievement gap with native-English speakers by Grade 11, after both groups have received at least five years of schooling in the U.S. When you examine the group means of former ELLs, LM students, and native-English speakers (see step 3 above), are these group means the same or within a five-NCE range? Are the means of former ELLs and LM students at or close to the 50th NCE (the mean of the national norm group)?

If the answer is "yes," then congratulations! Your existing school practices are allowing English language learners to achieve instructional parity with native-English speakers in a five-year period. This means that your instructional practices are very successful by stringent criteria, and you have passed the Thomas-Collier test that determines if English language learners have received full equal educational opportunity in your school system.

If the answer is "no," then more questions are in order. Is the achievement gap in Grade 11 smaller, the same size, or larger than it was when these students were last tested? If the answer is "larger," then your students are failing to make the "one year's progress in one year's time" that is necessary for them to keep up with native-English speakers. If the answer is "the same size," then your students have averaged "one year's progress in one year's time" for the past several years, thus maintaining the existing gap but not closing it. If the answer is "smaller," then your students have outgained the native-English speakers, but not by enough to allow them to close the achievement gap in the goal of five years.

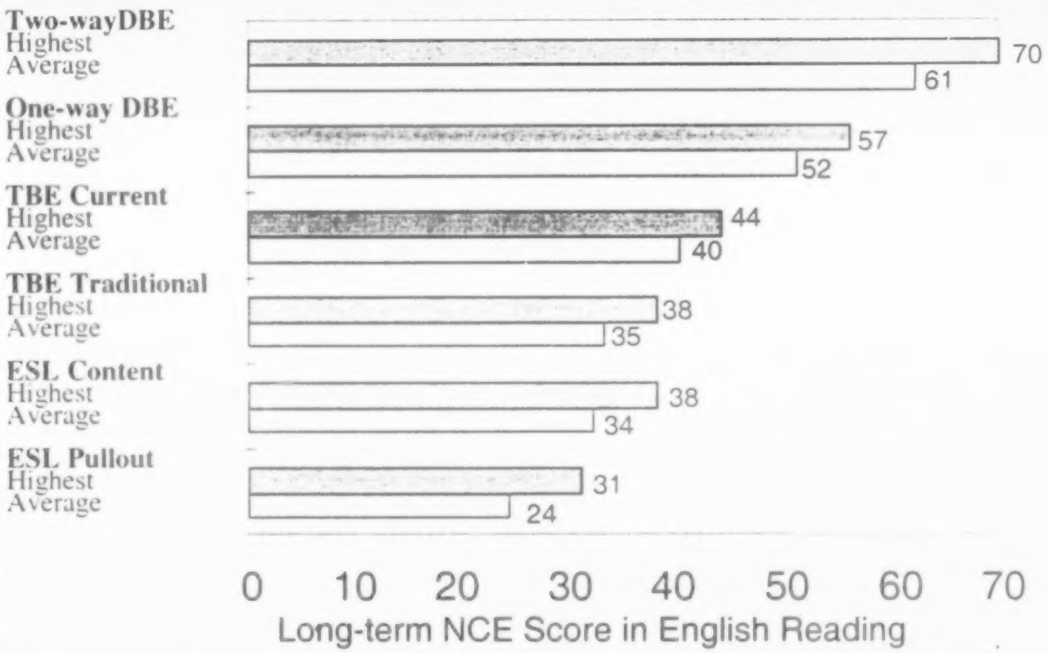
If You Failed the Thomas-Collier Test

We have examined test data and reviewed testing summaries from school systems in more than half of the states in the U.S. during the past ten years. **Based on this experience, we can say that a large majority of school systems have instructional practices for English language learners that cause them to fall short of passing the Thomas-Collier test.** Compare our findings, summarized in Figure 9, to your findings in your school district. If our findings match your school system's results, then it is appropriate for you to examine several additional factors:

(1) Are there good theoretical reasons to believe that your chosen instructional practices *should* be effective in allowing English language learners to reach eventual achievement parity with native-English speakers. In particular, do your instructional practices address

Figure 9

Average and Highest Group Long-term Achievement of Former English Learners in 11th Grade English Reading NCEs by Well-implemented Program Type



NOTE: Students began exposure to English in Kindergarten, attended one of the above programs in elementary school, and received regular mainstream instruction in English during middle and high school.

each of the components of Thomas and Collier's Prism Model of Language Acquisition for School?

(2) Is your school program well-implemented? Are your teachers well-trained in the instructional methods that "deliver" your chosen programs' impact? Do your principals actively support the classroom instruction in their schools? Have your programs stabilized and improved from their beginnings? If you have chosen instructional practices that are theoretically sound, but long-term results are less than expected, it is entirely appropriate to look for ways to improve the implementation of your present practices.

(3) It is also appropriate for you to seek instructional practices that have been shown to be effective in enhancing achievement gains by English language learners. In our research, the three major predictors discussed earlier contain several effective instructional practices that you might consider adding to your programs.

BEST COPY AVAI

Action Recommendations

Based on our collaborative data collection and data analysis, we have made the following recommendations to each of the school systems with whom we have jointly engaged in action-oriented research during the first five years of our current research program, from 1991-96. We pass on these recommendations to other school systems, based on our findings from our work with our five participating school systems and based on the findings of many other researchers in the field.

Action 1: Don't 'water down' instruction for English language learners and don't completely separate them from the instructional mainstream for many years, but also don't dump them into the mainstream unassisted until they are ready to successfully compete with native-English speakers when taught in English. English learners need on-grade-level instruction in their first language while they are learning English, the same cognitive development opportunities as native-English speakers receive, and continued assistance after they enter the regular instructional program.

Action 2: Provide opportunities for parents to assist their children using the parents' first language, the one they know best and the one in which they can best interact with their children at a higher cognitive level. Parents, even those with little education, can help you with their children's cognitive development at home. With help from you, they can assist in their children's academic development at home as well. Both of these can help prevent the cognitive and academic slowdown that can occur when students are taught exclusively in English at school. In this way, parents can provide the first language support that may be missing in the school and that helps English learners keep up with the native-English-speaking peers' rate of cognitive and academic progress while they are learning English. Parents can also provide a learning microcosm that is favorable toward their first language, thus giving their child the documented advantages of an additive bilingual environment, even if the school represents a subtractive environment.

Action 3: Provide continuing cognitive development and academic development while your students are learning English by means of the use of their first language in instruction for a part of each school day. They need to reach full development of their first language in order to fully develop their second language, English. Don't let them experience cognitive slowdown or academic slowdown, relative to the native-English speakers, while they are acquiring English to a level necessary to successfully compete with the native-English speakers on academic tasks and tests in English on grade level.

Action 4: Use current approaches to instruction, emphasizing interactive, discovery learning and raising the cognitive level of instruction in all classrooms by avoiding 'drill and kill' programs that may have positive short-term effects but which fail to allow students to sustain their achievement gains across time and to reach full parity with native speakers of English. Students working cooperatively together in a socioculturally supportive classroom do better than those taught traditionally. Provide ongoing staff development for teachers to share and co-develop cooperative learning, thematic lessons, L1 and L2 literacy development across the curriculum, process writing, performance and portfolio assessment, uses of technology, multiple intelligences, critical thinking, learning strategies, and global perspectives infused into the curriculum.

Action 5: Improve the sociocultural context of schooling for all of your students, English learners and native-English speakers alike. This means that your school should become an additive bilingual environment, viewing bilingualism as enrichment, even while your community may represent a highly subtractive language-learning environment. In a socioculturally supportive school, all students and staff and parents are respected and valued for the rich life experiences in other cultural contexts that they bring to the classroom. The school is a safe, secure environment for learning, and students treat each other with respect, with less expression of discrimination, prejudice, and hostility.

Action 6: If you can, try to move away from an emphasis on all-English instruction and move away from less effective forms of bilingual education. Try to move toward one-way and two-way developmental bilingual education (mainstream, enrichment bilingual education, rather than remedial approaches) as the program alternatives that may allow your students to eventually reach full educational parity with native speakers of English in your school.

Action 7: If, for pragmatic and practical reasons (e.g., a low-incidence language or shortage of bilingual teachers), you must use all-English instruction, select and develop its more effective forms. Specifically, try to move your school away from its least effective form, ESL pullout, and move toward the use of ESL taught through academic content and current approaches to teaching as a more efficacious alternative that helps students develop academically and cognitively to a greater degree. Develop your ESL-content program fully over the next 3-5 years by engaging your staff in professional development activities that increase their understanding of the theory and teaching practices associated with this program, so that you improve the degree to which it is fully and faithfully implemented. Look for alternatives that address students' cognitive needs as well--one example is the Cognitive Academic Language Learning Approach (CALLA; see Chamot & O'Malley, 1994).

Action 8: If you are now implementing transitional bilingual education (TBE) at elementary school level, try to move toward an alternative that is even more effective in the long-term--one-way or two-way developmental bilingual education. Although a well implemented TBE program is associated with significantly higher long-term achievement than ESL-content, neither program closes the achievement gap between English language learners and native-English speakers in the long-term.

Action 9: If you are now implementing two-way bilingual education, work on more fully developing a valid and effective implementation of this approach. Consider offering this program at the middle school, and later at the high school, for those students who were exposed to it in elementary school. At the middle and high school levels, merge this program with existing programs in foreign language for native-English speakers.

Action 10: If you're concerned about cost-effectiveness, be aware that it is most cost-effective to teach the grade-level, mainstream curriculum (not a watered-down version) to English language learners and language minority students who are proficient in English using a bilingual teacher, teaching a mainstream bilingual class. The costs of this approach are the same as in any class, except for the added cost of curricular materials in two languages. **ESL pullout is the least cost-effective model,** because extra resource teachers are needed.

Action 11: Think "enrichment" rather than "remediation" when you design programs for English language learners. Your English learners are not "broken" and they don't need fixing. What they do need is an opportunity to keep up in academic and cognitive development while they are enriching themselves by adding the world's most powerful language, English, to their own language. They have acquired their first language naturally from birth and have continued to develop this spoken language to age-appropriate level, providing them with a natural resource to assist our country, in the global community that exists today. What we all want is for English learners to be well educated and to learn English. The best and most effective way to accomplish this is to allow them to continue to develop their first language, and to use it to continue their cognitive and academic development, while they are learning English to a level commensurate with that of a native-English speaker. If they can end their schooling with good cognitive development, good academic development, a native-speaker command of English, as well as well-developed first language, so much the better! They will enrich themselves and our society by doing so.

Two languages are better than one--for English language learners and for native-English speakers alike. Learning two (or more) languages is the hallmark of the educated person, and is encouraged in the academic circles of the college-bound high school student and in higher education. Why not bring the enrichment advantages of learning two languages to a wider circle of students, including language minority students as well as native-English speakers?

A Call to Action

After you have heard the strongly-voiced opinions (most with little or no long-term data to support them) on both sides of this politically charged debate, you will need to make decisions. We offer you the same advice that we have offered to our collaborating school systems.

First, examine what the researchers have to say, but remember that you cannot be sure that they have answers that are completely meaningful for your local context. Second, listen to the political debate, but remember that the debaters' answers will gloss over, or even conspicuously ignore, the facts and established understandings that are inconvenient to their case. Most strongly held opinion in this field is motivated by emotions of nationalism or ethnic pride, fear that the world's most powerful language will be replaced in our country, fear of immigration or diversity, or fear of oppression by majority groups. Those who offer strong personal opinions often have little or no theoretical or professional understanding of the needs of language minority students. They frequently seek to support their politically-based opinions with poor understanding of available short-term research studies done on small groups that offer little real guidance to school-based decision-makers who need to know the long-term impact of their curricular choices. The solutions offered to complex questions by those on all sides who have strong personal opinions but weak or no data to support them are likely to be simplistic, and wrong.

Third, we advise that school personnel examine the results of their professional practice for the past 5-6 years by taking the Thomas-Collier test of equal educational opportunity as described above. This basic comparison of the past performance of English language learners, language minority students, and native-English speakers in your school system will offer much insight as to how well your present practices are working. We urge that you look at your own student data in this way either privately or publicly, depending on your local political context, but we most urgently advise that you *do* examine how your own students have fared. We hope that you find that the achievement gap between your English language learners and your native-English speakers has narrowed or closed during the past 5-6 years, but our experience with our five collaborating school systems, and with other school systems in 26 states with whom we have met and compared research results, leads us to predict that your findings will closely match our national findings as presented in this publication.

Fourth, when you have convinced yourself that there is a large achievement gap in your school district that needs to be addressed, and when you have reminded yourself that language minority students, now poorly served, are a "growth industry" nationally in education during the next 15-20 years (probably in your school system as well), you are ready to begin the process of constructive reform. We urge that you adopt the Prism Model as your construct for change, and that you seek to close the academic, cognitive, and linguistic gaps between your English language learners and your native-English speakers in all ways possible, in a socioculturally supportive environment for both groups, not just the native-English speakers. This will require careful study, a long-term commitment to constructive reform, and a willingness to exercise creative and effective professional leadership in your community based on knowledge and caring action, rather than on polemics. To quote Schiller, "The full mind is alone the clear, and truth dwells in the deeps." In our national interest, and in the interest of your own students and your own community, we urge you to fill your minds with pertinent professional knowledge and to seek the deeper educational truths that apply to your schools and school district.

ENDNOTES

1. Wherever possible, we have adopted the term "grade-level" classroom from Enright and McCloskey (1988), to replace the term "mainstream" or "regular" classroom, in the spirit of many professionals' concerns to use terms with fewer negative associations in our field. "Mainstream" is used by the field of special education to distinguish between mainstream classes that all students attend, in contrast to special education classes in which students might be placed for a short or long period of time when students have special needs that cannot be met in the mainstream classroom. This term has been adopted by our field, but in many contexts it is not an appropriate term. We use the term "mainstream" when we are contrasting separate bilingual/ESL classes that may or may not be on grade level, in comparison to the curriculum for native-English speakers. When we use the term "grade-level," it refers to classes in which students are performing age-appropriate academic tasks at the level of cognitive maturity for their age and grade level. Many bilingual and ESL classes are also grade-level classes.

2. The term developmental bilingual education was first introduced in the 1984 Title VII federal legislation, to emphasize the students' ongoing linguistic, cognitive, and academic developmental processes in both L1 and L2. In this report, we use this term to represent all enrichment models of bilingual schooling, including bilingual immersion, dual language, maintenance, and late-exit bilingual education. All of these models emphasize a focus on academic enrichment through both languages with L1 grade-level academic work provided through at least the end of elementary schooling (ideally Grades K-12).

   The distinction between one-way and two-way refers to the language groups served in a bilingual program (Stern, 1963). In one-way bilingual education, one language group is schooled bilingually. Two-way bilingual education is an integrated model in which speakers of each of two languages (e.g. Spanish speakers and English speakers) are placed together in a bilingual classroom to receive instruction across the curriculum through both of their two languages. Two-way is a grade-level, mainstream bilingual program, since native-English speakers are included, and the class receives age-appropriate schooling across the curriculum.

3. The theoretical concept of additive and subtractive bilingualism was first developed by Wallace Lambert (1975). These terms refer to the societal context in which bilingualism develops. In an additive bilingual context, students acquire a second language at no cost to continuing cognitive and linguistic development in their first language. An additive bilingual context, with time, can lead to age-appropriate proficiency in both L1 and L2. Proficient bilinguals outscore monolinguals on school tests. Thus an additive bilingual setting leads to positive cognitive effects for proficient bilinguals. Whereas, in a subtractive bilin-

gual setting, as students acquire L2, they gradually lose L1. For example, this may happen in situations where the L2 is prestigious and the L1 is perceived as low in status, in relation to the high-status language. In subtractive bilingual settings, students losing L1 tend to do less well in school as the cognitive complexity increases in the school curriculum. Transforming a school into an additive bilingual context can dramatically improve bilingual students' academic achievement.

Appendix A

Percentiles and Normal Curve Equivalents (NCEs)

Relative vs. Absolute Measures of Achievement

Throughout our research, we use NCEs instead of percentiles, following federal education regulations that specify the use of NCEs for comparing programs and student groups on norm-referenced achievement tests. Percentiles are similar to NCEs in several ways: they both range from 1-99 and they both have an average score of 50. Also, they both measure *relative* achievement when used in a pretest-to-posttest comparison; a student must make a full-year's progress to maintain his/her initial percentile score over a one-year period.

In contrast, a score that measures *absolute* achievement increases across time as the student's number of correct answers increases. Examples of absolute measures include scale scores and raw scores. Absolute scores tell us when students are making progress but they do not tell us whether the students are making enough progress to keep up with their peers as they advance from grade to grade through the school years. Thus, it is quite possible for a student to "make progress" every year but end up with very low scores by the end of the school years, when compared to his/her peers who may have outgained our student each and every year of school. It is even possible for our student to make "good progress for his/her situation" every year and still end up in the bottom one-tenth of eventual graduates. Clearly, we need both absolute measures, to tell us how much progress our student is making each year, as well as relative measures, to tell us whether our student's absolute progress is less than, the same, or more than the progress of his/her fellow students across the years of schooling.

NCEs and Percentiles as Relative Measures of Achievement

Let's examine NCEs in more detail. If our student fails to make "one-year's progress-in-one-year's time" (as defined by the performance of comparison students who have the same pretest percentile), our student's NCE score will fall between pretest and posttest. In other words, a student who initially scored at the 50th NCE or percentile in the spring of 1997 must make "one-year's-progress-in-one-year's time" to stay at the 50th NCE when tested a year later in the spring of 1998. Why? Because the entire group of comparison students (called the norm group in a norm-referenced test) has moved ahead in achievement during the year. This comparison group represents a "moving target" that is constantly advancing in tested achievement, and our student at the 50th NCE must make "one-year's progress-in-one-year's time" to keep up with his/her constantly advancing peers, and maintain his/her 50th NCE score. Thus, a year-to-year gain of zero NCEs means that our student has made a "year's-progress-in-a-year's-time." A year-to-year gain of zero NCEs does NOT mean that the student has made no progress at all--it means that he/she has made as much progress as the typical student who scored at the 50th NCE the year before.

Similarly, an NCE gain of zero between pretest and posttest means that the average student (or group of students) has made "a-year's-progress-in-a-year's-time." A gain of 1, 2, 3 or more NCEs means that the student has outgained his/her comparable peers by making *more* than typical amounts of progress (i.e., more than "one-year's-progress-in-one-year's-time") and has advanced his/her relative position in the distribution of comparison students. The NCE gain represents achievement gains over-and-above the achievement gains of typical, comparable students.

NCEs Explained

OK, we've seen that both percentiles and NCEs measure relative student achievement as compared to the achievement of constantly-advancing, similar students in the norm group and we've seen that NCEs are similar to percentiles in several ways, but exactly what is an NCE and how is an NCE different enough from a percentile to justify its preferential use?

Simply put, an NCE is a percentile that has been "transformed" to fix a serious problem of percentiles, the fact that percentiles are ranks and that the achievement "distance" between consecutive percentiles changes. So what's the problem? Well, if the range of scores in a normal distribution is divided into equal-sized standard deviations, the five percentile difference between percentiles 1 and 6 represents about three-fourths of a standard deviation. However, another 5 percentile difference, when it occurs between percentiles 45 and 50, represents only about one-eighth of a standard deviation. This means that a 5 percentile difference is a different amount of achievement depending on how high or low the percentile value is! Percentiles are smaller in the middle of the normal distribution (where about 34 percentiles fit in one standard deviation) than they are in the extremes of the normal distribution (where about 2 percentiles fit in one standard deviation) precisely because there are more people (or test scores) clustered in the middle of the normal distribution than in the extremes.

In the above example, the achievement difference between percentiles 1-6 is about six times larger than the achievement difference between percentiles 45-50. In other words, the actual amount of achievement represented by one percentile (or five percentiles) changes as one moves across the possible percentile values of 1 to 99. Percentiles are really ranks (e.g., first, second, third) and the achievement difference between consecutive ranks changes as one moves up or down the ranks from 1-99. We experience this phenomenon of differing distances between ranks in the real world when we remember that the first place finisher in a race (rank 1) may finish one foot ahead of the second place finisher (rank 2) but 100 feet ahead of the third place finisher (rank 3). In this example, using percentiles is similar to using the 1-2-3 ranks. In contrast, using an interval score, such as NCEs, is similar to measuring the distance between finishers in feet, an equal-interval unit of measurement.

Another way to conceptualize the use of percentiles is to imagine trying to measure distance with a yardstick that was constantly changing in size (thus changing the definition of a yard from 36 inches to some other value) as one used it. This would create assessment havoc in interpreting measurements. In similar fashion, educators who add and subtract percentile scores when comparing the test scores of groups or programs run a grave risk of making fundamental errors of interpretation and of making incorrect decisions for their students when using these test scores for decision-making. So, what we need is a test score with characteristics that many educators mistakenly believe are characteristics of percentiles. Specifically, we need a percentile-like score (with values 1-99 and average score of 50) *whose units remain the same size from values 1 through 99.*

How NCEs Are Computed

Testing companies who provide national NCEs in test reports have already done the statistical transformation of percentiles to NCEs for you, so you do not need to do this yourself. But if you wanted to convert percentiles to NCEs yourself, you could do it as follows:

(1) look up each percentile from 1-99 in a z-score table from a normal distribution, and write down the z-score (fraction of a standard deviation above or below the mean) that corresponds to each percentile. For example, a percentile of 9 corresponds to a z-score of -1.34, indicating that this score is 1.34 standard deviations below the mean of the distribution.

(2) take each z-score (remembering that z-scores represent interval data since standard deviations are the same size across the normal distribution), multiply it by 21.06 and add 50. In our example, a z-score of -1.34 is equivalent to an NCE of 21.8 or about 22.

Why do this? Because the result is a distribution of equal-sized scores from 1 to 99 with a mean of 50 and a standard deviation of 21.06. Why did we choose 21.06 as the standard deviation? Because that's what it takes to get NCE scores that range from 1 to 99, imitating percentiles. Thus, NCEs are "transformed percentiles" in that they represent percentiles that have been statistically (and legitimately so!) transformed so that the new "converted percentiles" have values from 1-99 (just like percentiles), have a mean of 50 (just like percentiles), but are equal in size across the 1-99 range of scores (unlike percentiles). Perhaps the best description of NCEs is that NCEs are what many educators have always believed that percentiles were, but they were wrong!

How NCEs Are Used

Now we can add, subtract, and compare equal-sized scores from different students, from different schools, from different instructional programs, and even from different norm-referenced tests, as long as they were normed on well-selected national random samples of students, and were normed close together in time so that the random samples from each test are from the same national population of students. This is important since the characteristics of the national population and their performance on test items can change over a decade or so, requiring a re-norming of the test usually to make it more difficult as students master curricular material at earlier and earlier ages. Since scores from norm-referenced tests that meet these criteria are based on the standards of the normal distribution, an unchanging mathematical construct, the scores from these different norm-referenced tests can be compared defensibly, at least when they are from similar time periods.

Programs that produce student achievement gains of 5 NCEs are producing gains that are equivalent to about one-fourth (25%) of a standard deviation (5 NCEs divided by 21.06 NCEs in a standard deviation). Thus, a 5 NCE gain is one-fourth of a standard deviation more than the expected gain of zero NCEs. Many program evaluators consider gains of one-fourth of a standard deviation or higher to be both statistically significant and practically significant (i.e., worthy of use in "real-world" decision-making), even for small groups of students, such as the 25 students in a typical classroom.

Standards for Effective Instructional Programs Using NCEs

In our research, we look diligently for instructional programs that not only produce student achievement gains of 4-6 NCEs in one school year, but continue to do so, year after year. Why? First, because typical English language learners in such a program will be able to close the initial 25-30 NCE achievement gap with native-English speakers in about 5-6 years, if they demonstrate sustained NCE gains of 5 NCEs per year for 5-6 consecutive years. Second, an instructional program that consistently produces student achievement gains of 5 NCEs is an unusually effective program. Typical instructional programs for English language learners allow these students to make gains of 0 NCEs (one full year's progress as compared to the progress of the typical native-English speaker) to 3 NCEs (a gain of about one-seventh of a standard deviation more than the typical native-English speaker). Programs of moderate-to-strong effectiveness allow typical participating students to gain from 4-6 NCEs (about one-fifth to one-fourth of a national standard deviation) per year more than the "comparison group," the national sample of mostly native-English speakers from the test's norm group. Programs of outstanding and extraordinary effectiveness allow their average students to gain from 7-9 NCEs (equivalent to one-third to one-half of a national standard deviation) per year more than the native-English speakers.

Programs that show an average annual student gain of 10 NCEs or more are somewhat suspect and require additional examination in that apparent gains of this size are typically caused by factors other than legitimate program effects. Such large gains can be produced when small groups are examined, since the standard error of group means is much larger for small groups (e.g., 10-25 students) than for large groups (more than 100 students). Also, some tests from small test companies have ill-constructed norms and poor (or non-existent) random samples of the national student population, both of which can lead to NCE gains that are artificially "inflated." Finally, gains above 10 NCEs per year can be incorrectly produced by accidental errors in test scoring or by outright fraud. An example from our experience of many years ago is provided by a good friend who used the same pre-test norms to score his September testing as well as his testing of the following spring. By using pre-test norms to score his students' post-tests, he used the fall "standards" to evaluate his students' spring performance of nine months later, thus in effect artificially adding almost a year's achievement to each student's score, arriving at "gains" of 15 NCEs. A bit of gentle probing and explanation convinced our friend of his potentially embarrassing mistake and we were able to revise his scores to their true levels before his error became known.

For language minority students, the moral of the above story is that large-group gains of more than 10 NCEs per year are highly suspect and most unlikely to be real. A true group gain of 10 NCEs means that the group has gained an amount equal to a full year of instruction plus outgaining the national comparison group by an additional one-half of a national standard deviation. In 25 years of evaluation experience, the authors have never seen legitimate program gains of this magnitude for student groups of any size (more than 100), but we have occasionally seen such gains for small groups (5-10 students) and often for individual students. These gains occur because the uncertainty in measurement is much greater for individuals or small groups than for large groups, leading to occasional gains that are spurious and that are not sustained across time. For large groups of students, such gains are part of our optimistic, hopeful, and wishful thinking as educators who want to help English learners, but we have found that legitimate annual gains of more than 10 NCEs are virtually non-existent (and perhaps verging on impossible) for large groups of students in the "real world".

This tells us that those who assert that typical English-learners can reach full parity with native-English speakers in 1-2 years are fantasizing, since this would require the typical 25-30 NCE achievement gap between these groups to be closed at the rate of 15-30 NCEs per year. It just doesn't happen that way for large groups of students, although a rare individual student might demonstrate this level of progress for a year or two, especially if he or she were inappropriately administered a test in English before mastering enough English to fully understand the test items. In this case, the student's measured pre-test scores would underestimate his/her true performance in the short term, causing short-term, spuriously large pre-post gains as a result of artificially low pre-test scores. These falsely low pre-test scores could occur because the student couldn't fully understand the test content at pre-test but could do so by the time of the post-test at the end of the school year. After this short-term phenomenon has disappeared, average participating students will require 5-6 years to close the achievement gap, at average gains of 5-6 NCEs per year, the typical rate of gain for a strong program. **If there is to be a standard to which *all* programs for English learners should aspire, it is this:**

> **All well-implemented, strong programs for English learners should allow the average participating student to reach full educational parity with native-English speakers on all school subjects, tested on grade-level and in English, after 5-6 years of exposure to the instructional program, by allowing the participating students to gain at least 5 NCEs per year, for 5-6 consecutive years. After parity is achieved, the school program should allow typical English learners, who are now proficient in English, to show at least the same rate of achievement gain as native-English speakers (0 NCEs or more) until the end of their school years.**

APPENDIX B

PHASE II OF THOMAS AND COLLIER RESEARCH, 1996-2001:

A National Study of School Effectiveness for Language-Minority Students' Long-Term Academic Achievement

This research investigates long-term patterns in language minority (LM) students' academic achievement and student, program, and instructional variables that influence academic success in Grades K-12. The research consists of a series of studies conducted as collaborative research with the bilingual/ESL school staff in at least 10 school districts across the U.S. Research sites include school districts that have large numbers of language minority students, maintain well collected long-term data on these students, and provide many services for them, including two-way developmental bilingual education (90-10 and 50-50 models, K-5, K-8, and K-12), one-way developmental bilingual education (K-5, K-8), transitional bilingual education (L1 support for K-2 or K-3), ESL taught through academic content (K-12), and ESL pullout or ESL as a subject (K-12). Collaborative, policy and decision-oriented analyses of the data are conducted with school staff, and interpretation considers the sociocultural contexts in which the students function.

Major Research Questions: Three major research questions frame the study:

- What are the characteristics of LM students in terms of their primary language, country of origin, L1 and L2 proficiency, prior academic performance, school attendance, degree of student retention in grade, socioeconomic status, and other student background variables?
- How much time is required for LM students to become academically successful after participating in the various bilingual/ESL programs, characterized as stable, well-established, and well-operated?
- What are the most important student, program, and instructional variables that affect the school achievement of LM students?

Study Design:

The researchers are collecting data from a variety of sources within each participating school system, including records from testing offices, centralized student information systems, LM central registration centers, and surveys of teachers, students, and parents conducted by participating school systems. School staff are being interviewed to collect information on the sociocultural context of schooling within each instructional setting.

The researchers use data capture software and relational database computer programs to take data from these sources and restructure them into a comprehensive LM student database for each school system. Analyses include descriptive summaries for each variable, as well as exploratory data plots and graphical analyses. Hierarchical multiple linear regression is used to explore the relative importance of student, program, and sociocultural variables on long-term student outcomes.

Analyses for each individual school district are provided as internal reports to each school district. The national research reports from these analyses will focus on analysis of general patterns in the data across multiple school district sites.

The study includes analysis of background factors that may influence LM student academic achievement, such as amount of English proficiency, poverty, geographic location, country of origin or ethnicity, and amount of prior schooling. Subjects for this study include U.S.-born and immigrant populations of Hispanic, Asian, and other LM background, including over 100 different language groups, as well as American Indian groups.

From 1996-2001, Phase II of this study is being funded by the Office of Educational Research and Improvement of the U.S. Department of Education. This study is one of 30 studies being conducted under the auspices of the Center for Research on Education, Diversity, and Excellence (CREDE), located at the University of California, Santa Cruz. The director of the CREDE Center is Dr. Roland Tharp. CREDE publications over the next five years will summarize the findings from these 30 studies and the implications for education practitioners.

REFERENCES

August, D., & Hakuta, K. (Eds.). (1997). *Improving schooling for language-minority children: A research agenda.* Washington, DC: National Academy Press.

August, D., & Pease-Alvarez, L. (1996). *Attributes of effective programs and classrooms serving English language learners.* Santa Cruz, CA: National Center for Research on Cultural Diversity and Second Language Learning.

Baker, C. (1993). *Foundations of bilingual education and bilingualism.* Clevedon, England: Multilingual Matters.

Baker, K., & de Kanter, A.A. (1981). *Effectiveness of bilingual education: A review of the literature.* Washington, DC: U.S. Department of Education.

Berko Gleason, J. (1993). *The development of language* (3rd ed.). New York: Macmillan.

Berliner, D.C., & Biddle, B.J. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools.* Reading, MA: Addison-Wesley.

Bialystok, E. (Ed.). (1991). *Language processing in bilingual children.* Cambridge: Cambridge University Press.

Branigin, W. (1996, March 23). Unusual alliance transformed immigration debate: Wide variety of interests opposed Congressional effort to limit influx of legal aliens. *Washington Post,* p. A8.

Castañeda v. *Pickard*, 648 F.2d 989 (5th Cir. 1981).

Chamot, A.U., & O'Malley, J.M. (1994). *The CALLA handbook: Implementing the Cognitive Academic Language Learning Approach.* Reading, MA: Addison-Wesley.

Chu, H.S. (1981). *Testing instruments for reading skills: English and Korean (Grades 1-3).* Fairfax, VA: Center for Bilingual/Multicultural/ESL Education, George Mason University.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum.

Collier, V.P. (1987). Age and rate of acquisition of second language for academic purposes. *TESOL Quarterly, 21,* 617-641.

Collier, V.P. (1988). *The effect of age on acquisition of a second language for school.* Washington, DC: National Clearinghouse for Bilingual Education.

Collier, V.P. (1989). How long? A synthesis of research on academic achievement in second language. *TESOL Quarterly, 23*, 509-531.

Collier, V.P. (1992). A synthesis of studies examining long-term language minority student data on academic achievement. *Bilingual Research Journal, 16*(1-2), 187-212.

Collier, V.P. (1995a). *Acquiring a second language for school.* Washington, DC: National Clearinghouse for Bilingual Education.

Collier, V.P. (1995b). *Promoting academic success for ESL students: Understanding second language acquisition for school.* Elizabeth, NJ: New Jersey Teachers of English to Speakers of Other Languages-Bilingual Educators.

Collier, V.P. (1995c). Second language acquisition for school: Academic, cognitive, sociocultural, and linguistic processes. In J.E. Alatis et al. (Eds.), *Georgetown University Round Table on Languages and Linguistics 1995* (pp. 311-327). Washington, DC: Georgetown University Press.

Collier, V.P., & Thomas, W.P. (1989). How quickly can immigrants become proficient in school English? *Journal of Educational Issues of Language Minority Students, 5*, 26-38.

Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation design and analysis issues for field settings.* Chicago: Rand McNally.

Cummins, J. (1976). The influence of bilingualism on cognitive growth: A synthesis of research findings and explanatory hypotheses. *Working Papers on Bilingualism, 9*, 1-43.

Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics, 11*(2), 132-149.

Cummins, J. (1991). Interdependence of first- and second-language proficiency in bilingual children. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 70-89). Cambridge: Cambridge University Press.

Cummins, J. (1996). *Negotiating identities: Education for empowerment in a diverse society.* Los Angeles, CA: California Association for Bilingual Education.

Cummins, J., & Swain, M. (1986). *Bilingualism in education.* New York: Longman.

Díaz, R.M., & Klingler, C. (1991). Towards an explanatory model of the interaction between bilingualism and cognitive development. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 167-192). Cambridge: Cambridge University Press.

Dolson, D.P., & Lindholm, K.J. (1995). World class education for children in California: A comparison of the two-way bilingual immersion and European school models. In T. Skutnabb-Kangas (Ed.), *Multilingualism for all*. Lisse, The Netherlands: Swets & Zeitlinger.

Dulay, H., & Burt, M. (1980). The relative proficiency of limited English proficient students. In J.E. Alatis (Ed.), *Current issues in bilingual education* (pp. 181-200). Washington, DC: Georgetown University Press.

Duncan, S.E., & De Avila, E.A. (1979). Bilingualism and cognition: Some recent findings. *NABE Journal, 4*(1), 15-20.

Enright, D.S., & McCloskey, M.L. (1988). *Integrating English: Developing English language and literacy in the multilingual classroom*. Reading, MA: Addison-Wesley.

Freeman, Y.S., & Freeman, D.E. (1992). *Whole language for second language learners*. Portsmouth, NH: Heinemann.

García, E. (1993). Language, culture, and education. In L. Darling-Hammond (Ed.), *Review of research in education* (Vol. 19, pp. 51-98). Washington, DC: American Educational Research Association.

García, E. (1994). *Understanding and meeting the challenge of student cultural diversity*. Boston: Houghton Mifflin.

Gardner, H. (1993). *Multiple intelligences: The theory in practice*. New York: Basic.

Genesee, F. (1987). *Learning through two languages: Studies of immersion and bilingual education*. New York: Newbury House.

Genesee, F. (Ed.). (1994). *Educating second language children: The whole child, the whole curriculum, the whole community*. Cambridge: Cambridge University Press.

Gonick, L., & Smith, W. (1993). *The cartoon guide to statistics*. New York: HarperCollins.

Hakuta, K. (1986). *Mirror of language: The debate on bilingualism*. New York: Basic Books.

Lambert, W.E. (1975). Culture and language as factors in learning and education. In A. Wolfgang (Ed.), *Education of immigrant students*. Toronto: Ontario Institute for Studies in Education.

Lau v. *Nichols*, 414 U.S. 563 (1974).

Lessow-Hurley, J. (1990). *The foundations of dual language instruction*. New York: Longman.

Light, R.J., & Pillemer, D.B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

Lindholm, K.J. (1990). Bilingual immersion education: Criteria for program development. In A.M. Padilla, H.H. Fairchild & C.M. Valadez (Eds.), *Bilingual education: Issues and strategies* (pp. 91-105). Newbury Park, CA: Sage.

Lindholm, K.J. (1991). Theoretical assumptions and empirical evidence for academic achievement in two languages. *Hispanic Journal of Behavioral Sciences, 13*, 3-17.

Lindholm, K.J. & Aclan, Z. (1991). Bilingual proficiency as a bridge to academic achievement: Results from bilingual/immersion programs. *Journal of Education, 173*, 99-113.

Lindholm, K.J., & Molina, R. (in press). Learning in dual language education classrooms in the U.S.: Implementation and evaluation outcomes. In *Proceedings of the III European Conference on Immersion Programmes*.

Lucas, T., Henze, R., & Donato, R. (1990). Promoting the success of latino language-minority students: An exploratory study of six high schools. *Harvard Educational Review*, 60, 315-340.

McLaughlin, B. (1992). *Myths and misconceptions about second language learning: What every teacher needs to unlearn*. Santa Cruz, CA: National Center for Research on Cultural Diversity and Second Language Learning.

McLeod, B. (1996). *School reform and student diversity: Exemplary schooling for language minority students*. Washington, DC: National Clearinghouse for Bilingual Education.

Moll, L.C., Vélez-Ibáñez, C., Greenberg, J., & Rivera, C. (1990). *Community knowledge and classroom practice: Combining resources for literacy instruction*. Arlington, VA: Development Associates.

National Education Goals Panel. (1994). *The national education goals report 1994: Building a nation of learners*. Washington, DC: U.S. Government Printing Office.

Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven: Yale University Press.

Oakes, J. (1992). Can tracking research inform practice? Technical, normative, and political considerations. *Educational Researcher, 21*(4), 12-21.

Oakes, J., Wells, A.S., Yonezawa, S., & Ray, K. (1997). Equity lessons from detracking schools. In A. Hargreaves (Ed.), *Rethinking educational change with heart and mind* (pp. 43-72). Alexandria, VA: Association for Supervision and Curriculum Development.

Pedhazur, E.J. (1982). *Multiple regression in behavioral research: Explanation and prediction* (2nd ed.). Fort Worth, TX: Holt, Rinehart and Winston.

Pérez, B., & Torres-Guzmán, M.E. (1996). *Learning in two worlds: An integrated Spanish/English biliteracy approach* (2nd ed.). White Plains, NY: Longman.

Rossell, C.H., & Baker, K. (1996). The educational effectiveness of bilingual education. *Research in the Teaching of English, 30*(1), 7-74.

Simon, J. (1993). *Resampling: The new statistics. Boston: Wadsworth.*

Skutnabb-Kangas, T. (1981). *Bilingualism or not: The education of minorities*. Philadelphia: Multilingual Matters.

Snow, C.E. (1990). Rationales for native language instruction: Evidence from research. In A.M. Padilla, H.H. Fairchild, & C.M. Valadez (Eds.), *Bilingual education: Issues and strategies*. Newbury Park, CA: Sage.

Stern, H.H. (Ed.). (1963). *Foreign languages in primary education: The teaching of foreign or second languages to younger children*. Hamburg: International Studies in Education, UNESCO Institute for Education.

Tabachnick, B.G., & Fidell, L.S. (1989). *Using multivariate statistics* (2nd ed.). New York: Harper & Row.

Tharp, R.G., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning, and schooling in social context*. Cambridge: Cambridge University Press.

Thomas, W.P. (1992). An analysis of the research methodology of the Ramírez study. *Bilingual Research Journal, 16*(1-2), 213-245.

Thomas, W.P., & Collier, V.P. (1996). Language-minority student achievement and program effectiveness. *NABE News, 19*(6), 33-35.

Thonis, E. (1994). Reading instruction for language minority students. In C.F. Leyba (Ed.), *Schooling and language minority students* (2nd ed., pp. 165-202). Los Angeles: Evaluation, Dissemination and Assessment Center, California State University, Los Angeles.

Tinajero, J.V., & Ada, A.F. (Eds.). (1993). *The power of two languages: Literacy and biliteracy for Spanish-speaking students.* New York: Macmillan/McGraw-Hill.

Willig, A.C. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research, 55*, 269-317.

Wong Fillmore, L. (1991). Second language learning in children: A model of language learning in social context. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 49-69). Cambridge: Cambridge University Press.

Wong Fillmore, L., & Valadez, C. (1986). Teaching bilingual learners. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 648-685). New York: Macmillan.

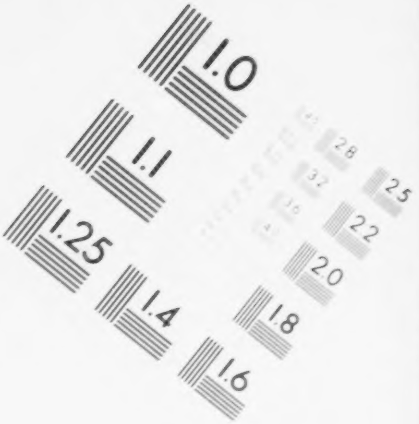
Zappert, L.T., & Cruz, B.R. (1977). *Bilingual education: An appraisal of empirical research.* Berkeley, CA: Bahía Press.

1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2

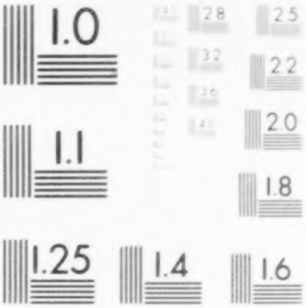
AIIM

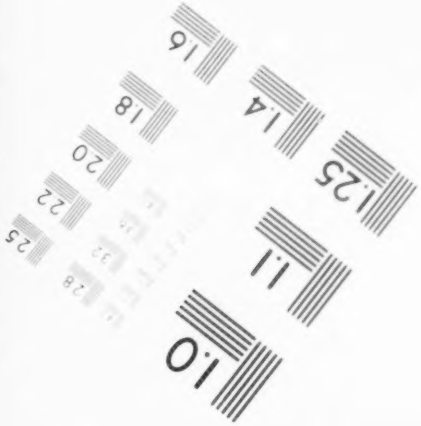
Association for Information and Image Management

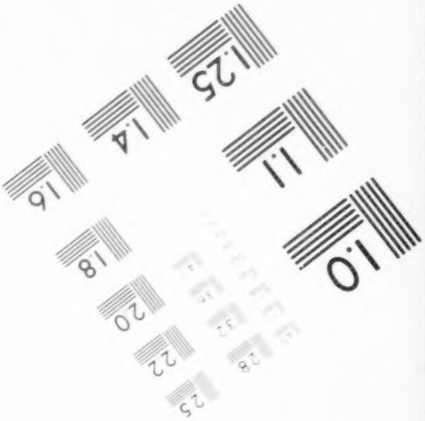
1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910
301/587-8202


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2

Centimeter
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
mm
1
2
3
4
5
Inches


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2

MANUFACTURED TO AIIM STANDARDS
BY APPLIED IMAGE, INC.

ABOUT THE AUTHORS

Dr. Wayne P. Thomas is a professor of Evaluation and Research Methodology in the Graduate School of Education at George Mason University in Fairfax, Virginia. His Ph.D. training and primary professional experience are in program evaluation methodology and social science research methods. He also has extensive experience in designing large-scale databases and developing computer software for purposes of student testing, program evaluation, and educational data management. His research and publications focus on the evaluation of school effectiveness for language minority and Title I students. He is a former computer programmer & analyst, high school math & physics teacher, and school system central office administrator in school planning, testing, and program evaluation. He teaches doctoral courses and advises doctoral students in program evaluation and quantitative research methods, and directs doctoral dissertations.

Dr. Virginia P. Collier is a professor of Bilingual/Multicultural/ESL Education in the Graduate School of Education at George Mason University. She is co-author with Carlos Ovando of the book, *Bilingual and ESL Classrooms: Teaching in Multicultural Contexts*, with the new second edition published by McGraw-Hill in November, 1997. This book is a well known, comprehensive reference on research, policy, and effective practices in the U.S. for students of culturally and linguistically diverse backgrounds. In addition, Dr. Collier has over 30 other publications in the field of language minority education, including her popular monograph, *Promoting Academic Success for ESL Students*, published by the New Jersey TESOL-BE professional association. She has served the field of language minority education for three decades as parent, teacher, teacher educator, and doctoral mentor.

During the past ten years, Drs. Thomas and Collier have collaborated in research on school effectiveness for linguistically and culturally diverse students. Their award-winning joint research has been utilized by many school systems in the U.S. and abroad to reform the education of language minority students and to promote school improvement for both native-English speakers and English language learners. Drs. Thomas and Collier have served as keynote speakers at many national and international conferences and have conducted education leadership training for superintendents, principals, and education policy makers in 26 U.S. states and 11 countries.